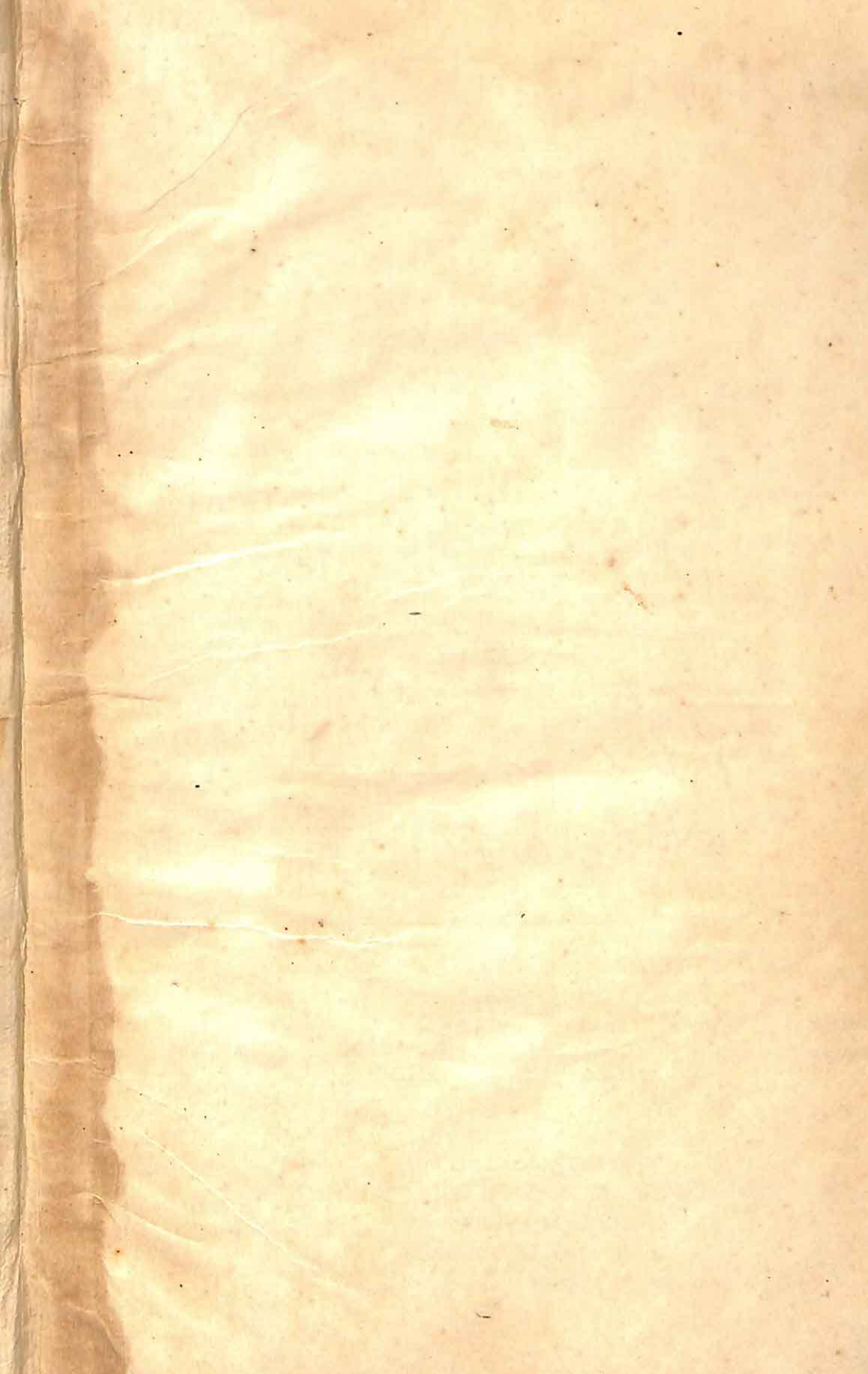






265  
19.70







# Psychological Bulletin

HARRY HELSON, Editor  
*Kansas State University*

## Consulting Editors

W. BEVAN, JR.  
*Kansas State University*

R. R. BLAKE  
*University of Texas*

W. R. GARNER  
*Johns Hopkins University*

J. P. GUILFORD  
*University of Southern California*

W. H. HOLTZMAN  
*University of Texas*

Q. MCNEMAR  
*Stanford University*

M. E. NOBLE  
*Kansas State University*

L. J. POSTMAN  
*University of California, Berkeley*

J. B. ROTTER  
*University of Connecticut*

S. B. SELLS  
*Texas Christian University*

W. A. WILSON, JR.  
*Bryn Mawr College*

---

## VOLUME 61, 1964

---

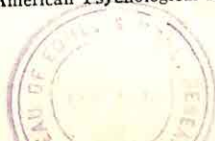
HELEN ORR  
*Managing Editor*

ELIZABETH S. REED  
*Advertising Manager*

VIRGINIA RICHARDS  
*Editorial Assistant*

PUBLISHED MONTHLY BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 SIXTEENTH ST. N. W.  
WASHINGTON, D. C. 20036

© 1964 by the American Psychological Association, Inc.



Bureau Edm. Res.

DAVID HARRIS

70 2 70

No 8 - 365

# CONTENTS OF VOLUME 61

BAHRICK, HARRY P. Retention Curves: Facts or Artifacts?	188
BAUMEISTER, ALFRED, HAWKINS, WILLIAM F., AND CROMWELL, RUE L. Need States and Activity Level	438
BEVAN, WILLIAM. Contemporary Problems in Adaptation-Level Theory: A Symposium	161
BEVAN, WILLIAM. Subliminal Stimulation: A Pervasive Problem for Psychology	81
BLACKWELL, H. RICHARD. Discussion: Competing Theories of Receptor Excitation	268
BLOCK, WILLIAM E. A Conceptual Framework for the Clinical Test Situation	168
BOYNTON, R. M. Discussion: Competing Theories of Receptor Excitation	262
BROWNFIELD, CHARLES A. Deterioration and Facilitation Hypotheses in Sensory-Deprivation Research	304
BUCHWALD, NATHANIEL A. See GARCIA, JOHN.	
CARRIERO, NICHOLAS J. A Note on "On Simple Methods of Scoring Tracking Error"	303
CERBUS, GEORGE, AND NICHOLS, ROBERT C. Significance Levels of Individual Correlations Selected from Multivariate Studies	400
CHAPANIS, ALPHONSE. See CHAPANIS, NATALIA P.	
CHAPANIS, NATALIA P., AND CHAPANIS, ALPHONSE. Cognitive Dissonance: Five Years Later	1
CROMWELL, RUE L. See BAUMEISTER, ALFRED.	
EHRlich, ANNETTE. Neural Control of Feeding Behavior	100
ENOCH, JAY M. Physical Properties of the Retinal Receptor and Response of Retinal Receptors	242
EPSTEIN, WILLIAM. Experimental Investigations of the Genesis of Visual Space Perception	115
FALEK, ARTHUR. See JARVIK, LISSY F.	
FURCHTGOTT, ERNEST. Postscript to "Behavioral Effects of Ionizing Radiations: 1955-61"	236
GAITO, JOHN, AND ZAVALA, ALBERT. Neurochemistry and Learning	45
GALLISTEL, C. R. Electrical Self-Stimulation and Its Theoretical Implications	23
GARCIA, JOHN, AND BUCHWALD, NATHANIEL A. Note on Furchtgott's "Behavioral Effects of Ionizing Radiations: 1955-61"	234
GARDNER, G. GAIL. The Psychotherapeutic Relationship	426
GOLDFARB, JOYCE LEVIS. See GOLDSTONE, SANFORD.	
GOLDSTEIN, IRIS BALSCHAN. Role of Muscle Tension in Personality Theory	413
GOLDSTONE, SANFORD, AND GOLDFARB, JOYCE LEVIS. Adaptation Level, Personality Theory, and Psychopathology	176
GRIDGEMAN, N. T. Exact Multinomial Significance Tests: Note on a Paper by Alphonse Chapanis	239
GUHL, A. M. Psychophysiological Interrelations in the Social Behavior of Chickens	277
GUILFORD, J. P. Zero Correlations among Tests of Intellectual Abilities	401
HARBURG, ERNEST. See MCGINN, NOEL F.	
HAWKINS, WILLIAM F. See BAUMEISTER, ALFRED.	
HEILBRUN, ALFRED B., JR. Social-Learning Theory, Social Desirability, and the MMPI	377
HELSON, HARRY. Competing Theories of Receptor Excitation in the Retina: A Symposium	241
HUFF, FREDERICK W. Learning and Psychopathology	459
JARVIK, LISSY F., FALEK, ARTHUR, AND PIERSON, W. PHILIP. Down's Syndrome (Mongolism): The Heritable Aspects	388
JULIUS, STEVO. See MCGINN, NOEL F.	
KEPPEL, GEOFFREY. Verbal Learning in Children	63
LANA, ROBERT E. Three Theoretical Interpretations of Order Effects in Persuasive Communications	314
LEVINGER, GEORGE. Note on Need Complementarity in Marriage	153
LOEB, MICHEL. Psychophysical Correlates of Intratympanic Reflex Action	140
MCCONNELL, DAVID G. Chemical Theories	252
MCGINN, NOEL F., HARBURG, ERNEST, JULIUS, STEVO, AND MCLEOD, JACK M. Psychological Correlates of Blood Pressure	209
MCLEOD, JACK M. See MCGINN, NOEL F.	
MARTIN, IRENE. Adaptation	35
NICHOLS, ROBERT C. See CERBUS, GEORGE.	

OVERALL, JOHN E. Note on Multivariate Methods for Profile Analysis.....	195
OVERALL, JOHN E. Note on the Scientific Status of Factors.....	270
PARDUCCI, ALLEN. Sequential Effects in Judgment.....	163
PAYNE, M. CARR, JR. Color as an Independent Variable in Perceptual Research.....	199
PIERSON, W. PHILIP. See JARVIK, LISSY F.	
REESE, HAYNE W. Discrimination Learning Set in Rhesus Monkeys.....	321
RENNER, K. EDWARD. Delay of Reinforcement: A Historical Review.....	341
ROSENTHAL, ROBERT. Experimenter Outcome-Oriented and the Results of the Psychological Experiment .....	405
SLOVIC, PAUL. Assessment of Risk Taking Behavior.....	220
SPENCE, KENNETH W. Anxiety (Drive) Level and Performance in Eyelid Conditioning...	129
STEISEL, IRA M. Note on "Personality Variables and Response to Color".....	399
THARP, ROLAND G. Reply to Levinger's Note.....	158
TUNE, G. S. Response Preferences: A Review of Some Relevant Literature.....	286
WEISS, JONATHAN H. Phonetic Symbolism Re-examined.....	454
ZAVALA, ALBERT. See GAITO, JOHN.	
ZEIGLER, H. PHILIP. Displacement Activity and Motivational Theory: A Case Study in the History of Ethology.....	362

# Psychological Bulletin

## COGNITIVE DISSONANCE:

### FIVE YEARS LATER<sup>1</sup>

NATALIA P. CHAPANIS AND ALPHONSE CHAPANIS

*Johns Hopkins University*

This article reviews critically the experimental evidence in support of cognitive dissonance theory as applied to complex social events. The criticisms which can be made of this literature fall into 2 main classes. 1st, the experimental manipulations are usually so complex and the crucial variables so confounded that no valid conclusions can be drawn from the data. 2nd, a number of fundamental methodological inadequacies in the analysis of results—as, e.g., rejection of cases and faulty statistical analysis of the data—vitiate the findings. As a result, one can only say that the evidence adduced for cognitive dissonance theory is inconclusive. Suggestions are offered for the methodological improvement of studies in this area. The review concludes with the thesis that the most attractive feature of cognitive dissonance theory, its simplicity, is in actual fact a self-defeating limitation.

Social psychologists have been trying for many years to predict the conditions under which attitudes and opinions are changed. In general their attempts have not been conspicuously successful. One of the first major breakthroughs in this area came when Leon Festinger (1957) published his book on *A Theory of*

*Cognitive Dissonance*. In this book the author presented a simple conceptual scheme by which he could predict with precision the outcomes of certain social situations. To support his theory, Festinger marshaled data from an impressive variety of field and experimental studies. In addition, he and other workers have since then conducted a number of studies designed to test specific derivations of the theory. What can we say about all this literature?

Cognitive dissonance theory has already been reviewed by Bruner (1957), Asch (1958), Osgood (1960), and Zajonc (1960). These writers, however, have been primarily concerned with a

<sup>1</sup> This article is based on material prepared by the senior author while she was employed at the Tavistock Institute of Human Relations, London, England. It appeared originally as Tavistock Document No. 626, "Cognitive Dissonance: A Dissenting Voice," dated June 1961. That work was part of a research program financed by S. H. Benson, Limited, and carried out by the Tavistock Institute. The extensive expansion and revision of the original document undertaken with the collaboration of the junior author was supported in part by Contract Nonr 248(55) between the Office of Naval Research and the Johns Hopkins University. We are pleased to acknowledge the assistance provided us by both organizations. The views expressed in this article are, however, those of the authors. Neither S. H. Benson, Limited, nor the Office of Naval Research is responsible for any of the statements contained in it.

We are greatly indebted to the staff of the Tavistock Institute, and particularly to Frederick E. Emery, for many helpful comments, guidance, and encouragement throughout the preparation of the original document. Grateful appreciation is also extended to our many American colleagues—too numerous to name individually—who read various versions of this article and offered helpful comments and encouragement.

critical evaluation of the conceptual system employed in dissonance theory. And, whatever they might think of the theory, most workers (except perhaps Asch) have been impressed by the scope, relevance, and ingenuity of the experimental evidence gathered in its support.

There is an engaging simplicity about Festinger's dissonance formulations. No matter how complex the social situation, Festinger assumes that it is possible to represent the meaning which the situation has for an individual by a series of elementary *cognitions*—statements that an individual might make describing his "knowledge, opinions or beliefs" (Festinger, 1957, p. 3). Moreover, a simple inventory of a group of related cognitions is sufficient to reveal whether or not they are consistent. The theory assumes further that people prefer consistency among their cognitions and that they will initiate change in order to preserve this consistency. So far these ideas are not new. They had been promulgated as early as 1946 by Heider with his concept of balance and imbalance. The magic of Festinger's theory, however, seems to lie in the ease with which imponderably complex social situations are reduced to simple statements, most often just two such statements. This having been done, a simple inspection for rational consistency is enough to predict whether or not change will occur. Such uncomplicated rationality seems especially welcome after having been told for years that our attitudes and resulting behavior are strongly dependent on motivational, emotional, affective, and perceptual processes (e.g., Krech & Crutchfield, 1948; Rosenberg, 1960).

Five years have now elapsed since the publication of Festinger's book, and this seems to be an appropriate time to pause for a close look at the evidence in support of the theory. For no matter how

appealing a theory might be, in the final analysis it is the evidence that counts. This paper, therefore, will be concerned with a review of experiments on cognitive dissonance in humans from two points of view. First we shall consider whether an experimenter really did what he said he did. Then later we shall consider whether the experimenter really got the results he said he did.

#### CONTROVERSIAL EXPERIMENTAL MANIPULATIONS

As we all know, good experimental work always involves manipulating conditions in such a way that we may ascribe changes we observe in our dependent variables to the manipulations we carried out on the independent variables. In actual practice we rarely define these manipulations in careful operational terms. When a pellet drops into a cup in front of a hungry rat we call it a reward, or reinforcement; when a wire transmits an electric shock to a person we call it punishment, or stress; and so on. Moreover, we do not, in general, quarrel with our fellow experimenter's interpretation of the situation. After all, he was there, he ought to know what it was about. However, when we deal with experiments on cognitive dissonance we have a very special problem on our hands.

#### *Experimental Dissonance*

Simply stated, cognitive dissonance theory is concerned with what happens when the cognitions of a person are discrepant. The basic premise is that discrepant cognitions create tension which the individual strives to reduce by making his cognitions more consistent. This tension is called cognitive dissonance, and the drive towards consistency, dissonance reduction. "When two or more cognitive elements are psychologically inconsistent, dissonance is created. Dis-

sonance is defined as psychological tension having drive characteristics" so that when dissonance arises the individual attempts to reduce it (Zimbardo, 1960, p. 86).

For our purposes at the moment the most important thing to note about the theory is that dissonance is an intervening variable whose antecedents are the private internal cognitions of a person. To test a theory like this, it is up to the experimenter to create various degrees of dissonance by introducing various discrepant cognitions within an individual. Whenever contradictory statements or syllogisms or opinions are used, there is not likely to be much controversy about the fact that they must lead to discrepant internal cognitions, and so, by definition, to dissonance. Indeed, studies on cognitive dissonance of this type have yielded results which are well-established, clear-cut, and consistent. But for the experiments under review here, the situation is rarely as simple as this. The Festinger group is primarily concerned with applying their dissonance formulation to predict complex social events. In order to do this experimentally, they use elaborate instructions and intricate relationships between experimenter and *S* to introduce discrepant cognitions and therefore to produce dissonance. Under such conditions, how can we be sure that the experimental situation has been successful in creating dissonance and dissonance alone?

In the face of such difficulties, it is always a good policy to ask the *S* himself about the situation, either directly or indirectly. It should be possible, for instance, to find out how the *Ss* perceived each of the experimental manipulations. One could also determine whether *Ss* perceived the situation as conflictful and, if so, to what extent. This kind of information is crucial to the theory of cognitive dissonance because all its pre-

dictions are based on the assumption that a state of differing, incompatible cognitions has been produced within the *S*. Unfortunately, evidence of this kind from the *Ss* themselves is not always available in the studies under review here. As a result, it is up to the reader to decide whether the experimental manipulations had the effect which the authors claim.

The other side of the coin, equally important, is that we must also assure ourselves that the experimental manipulations did not at the same time produce other internal states or cognitions within the *S* which could contaminate or even account for the findings. In fact, certain "nonobvious" derivations of some of these experiments may perhaps become a little more obvious when the experiments are reinterpreted to take other factors into consideration.

It is worthwhile spending a few moments on these nonobvious derivations. If we disregard the intermediate steps and simply consider the independent and dependent variables, it is possible to describe the essential aspects of some of these derivations by saying that they follow a *pain principle*. Reduced to essentials, some of Festinger's derivations say that the more rewarding a situation, the more negative is the effect; and contrariwise, the more painful a situation, the more positive is the effect. This is clearly illustrated by the following quotation from Festinger (1961): "Rats and people come to love the things for which they have suffered." However, if we carefully examine the kinds of experiments which are supposed to test these derivations, we find that, in general, the situation contains both painful and rewarding conditions, but that the manipulation is interpreted in terms of only one of these. It should be apparent that if a situation is both rewarding and painful, and the dependent variable

shows a positive effect, it is not legitimate to attribute it solely to the painful variable, or vice versa. To use a statistician's terminology, the variables are confounded.

Our most general criticism, then, is that some dissonance experiments have been designed in such a way that it is impossible to draw any definite conclusions from them.

### *Examples*

The best way of illustrating these points is to describe an experimental procedure and then to analyze it from two points of view: Did the experimenter really produce the discrepant cognitions he said he did? Did the experimental manipulations produce other cognitions that could contaminate or account for his findings?

*Relief or Dissonance?* Let us take this experiment: College women volunteered to participate in a series of group discussions on the psychology of sex. They were seen individually by the experimenter before being allowed to join an "on-going group." Some of the girls were told they would have to pass an embarrassment test to see if they were tough enough to stand the group discussion. They were free to withdraw at this point, and one *S* did so. Girls in the severe embarrassment group had to read out loud in the presence of the male experimenter some vivid descriptions of sexual activity and a list of obscene sex words. Another group of girls—the mild embarrassment group—read some mild sexual material. All of these girls were told that they were successful in passing the embarrassment test. Each *S* then listened as a silent member to a simulated, supposedly on-going group discussion, which was actually a standard tape recording of a rather dull and banal discussion about the sexual behavior of animals. A control group listened only

to the simulated group discussion. All groups then made ratings about this discussion, its participants, and their own interest in future discussions. The ratings made by the severe embarrassment group were, on the average, somewhat more favorable than those made by the other two groups.

What was this experiment about? Was it to demonstrate the effect of feelings of relief when people discover that a task (the group discussion) is not as painfully embarrassing as the embarrassment test led them to believe? No. Was it to demonstrate the effect of success in a difficult test (passing the embarrassment test) on task evaluation? No. Was it to demonstrate the displacement of vicarious sexual pleasure from a discomfiting, but sexually arousing, situation to a more socially acceptable one? No. The experimenters called it "The Effect of Severity of Initiation on Liking for a Group" (Aronson & Mills, 1959); that is, the more painful the initiation, the more the *Ss* like the group. They predict the outcome for the severe embarrassment group in the following way: In successfully passing the embarrassment test these girls "held the cognition that they had undergone a painful experience" in order to join a group; the discussion, however, was so dull and uninteresting that they realized the unpleasant initiation procedure was not worth it. This produced dissonance since "negative cognitions about the discussion . . . were dissonant with the cognition that they had undergone a painful experience." One of the ways they could reduce this dissonance was by re-evaluating the group discussion as more interesting than it really was.

All this may be so, but in order to accept the authors' explanation we must be sure the girls really did hold these discrepant cognitions, and no others. We have to be sure, for instance, that

they felt no relief when they found the group discussion banal instead of embarrassing, that success in passing a difficult test (the embarrassment test) did not alter their evaluation of the task, that the sexual material did not evoke any vicarious pleasure or expectation of pleasure in the future, and that the group discussion was so dull that the girls would have regretted participating. There is no way of checking directly on the first three conditions, although other experimental evidence suggests that their effect is not negligible. However, to check on the fourth factor we have the data from the control group showing that the group discussion was, in fact, more interesting than not (it received an average rating of 10 on a 0-15 scale). It is, therefore, difficult to believe that the girls regretted participating. To sum up, since the design of this experiment does not exclude the possibility that pleasurable cognitions were introduced by the sequence of events, and since, in addition, the existence of "painful" cognitions was not demonstrated, we cannot accept the authors' interpretation without serious reservations.

It is interesting to speculate what would have happened if the girls had been "initiated" into the group by the use of a more generally accepted painful procedure, such as using electric shock. Somehow it seems doubtful that this group would appreciate the group discussion more than the control group, unless—and here is the crucial point—the conditions were so manipulated that Ss experienced a feeling of successful accomplishment in overcoming the painful obstacle. It seems to us that if there is anything to the relationship between severity of initiation and liking for the group, it lies in this feeling of successful accomplishment. The more severe the test, the stronger is the pleasurable feeling of suc-

cess in overcoming the obstacle. There is no need to postulate a drive due to dissonance if a *pleasure principle* can account for the results quite successfully.

The same feeling of successful accomplishment may, incidentally, be the relevant variable involved in some of the "effort" experiments done by the Festinger group (e.g., Cohen, 1959). It seems reasonable to expect that in such experiments the higher the degree of perceived effort, the greater the feeling of successful accomplishment in performing a task. Thus, *effort* would be confounded with *feeling of success*. Note, however, that success is pleasant, whereas effort is painful. Here is a situation which could be both rewarding and painful, but dissonance workers see it only as painful. (Two other effort experiments will be analyzed in greater detail later in this section.)

*Reward or Incredulity?* Let us look at another experiment, this time by Festinger and Carlsmith (1959): Out of several possibilities, Ss chose to take part in a 2-hour experiment falsely labeled as an experiment on "measures of performance." The Ss were tested individually and were given a "very boring" and repetitive task for about 1 hour. At the end of the hour each S was given a false explanation about the purpose of the experiment. He was told that it was an experiment to test the effect of expectation on task performance. Some Ss were then asked if they would mind acting in a deception for the next couple of minutes since the person regularly employed for this was away. The Ss of one group were hired for \$1.00 each, those of another group for \$20.00 each, to tell the next incoming S how enjoyable and interesting the experiment had been (ostensibly the expectation variable). Each S was also told he might be called on to do this again. Some Ss refused to be hired. A control group of

Ss was not asked to take part in any deception. Subsequently, all Ss (control and hired Ss) were seen by a neutral interviewer, supposedly as part of the psychology department's program of evaluating experiments. During the interview, Ss were asked to rate the experiment along four dimensions. The only significant difference between the three groups was in terms of enjoyment. The control group rated the experiment as just a little on the dull side, the \$1.00 group thought it was somewhat enjoyable, and the \$20.00 group was neutral. The mean ratings for the control and \$20.00 groups were not significantly different from each other nor from the neutral point.

What was this experiment about? The authors call it "Cognitive Consequences of Forced Compliance." They make the prediction that "the larger the reward given to the S" the smaller the dissonance and therefore "the smaller the subsequent opinion change," and "furthermore . . . the observed opinion change should be greatest when the pressure used to elicit the overt behavior is just sufficient to do it." As an aside we should point out that, inasmuch as these statements clearly refer to a maximum and so by inference to some sort of a curvilinear or nonmonotonic relationship, it would have been better if more reward categories had been used. In addition, two more control groups—a *deception-but-no-reward* group, and a *reward-but-no-deception* group—should have been included to separate out the effects of reward and deception. However, our primary concern at the moment is not with such technical matters of experimental design.

Let us examine instead the meaning of the descriptive term "forced compliance." According to Festinger (1957), it means "public compliance without private acceptance [p. 87]."

The reward Ss, it is true, complied publicly with the instructions in that they described a boring task as enjoyable to another S. Notice, however, that even the control group rated the task as only slightly boring. This suggests that the false explanation placed the task in a wider context and may have led to "private acceptance" of the whole situation by both control and reward Ss. We could also question the choice of the word "forced." Forced implies a lack of freedom, but it is extremely difficult to predict how an S perceives his freedom of choice even when this variable is experimentally manipulated (e.g., Brehm & Cohen, 1959a). All we can say is that the term forced compliance is not a good description of the events in this experiment.

What seems to be even more important, however, is that the experiment could be more appropriately entitled "The Effect of a Plausible and Implausible Reward on Task Evaluation." As far as we can tell, Ss were not asked to describe their reactions to the size of the reward. Nevertheless, \$20.00 is a lot of money for an undergraduate even when it represents a whole day's work. When it is offered for something that must be much less than 30 minutes work, it is difficult to imagine a student accepting the money without becoming wary and alert to possible tricks. In fact, more than 16% of the original Ss in the \$20.00 group had to be discarded because they voiced suspicions, or refused to be hired. Under such circumstances, it seems likely that those who were retained might have hedged or been evasive about their evaluation of the experiment. The mean rating for the \$20.00 group was  $-.05$  on a scale that ranged from  $-5.00$  (dislike) to  $5.00$  (like), that is, the mean rating was at the neutral point. As other workers (e.g., Edwards, 1946) have suggested,

a rating at the zero or neutral point may be ambiguous, ambivalent, or indifferent in meaning and may simply represent an evasion. The authors' data, unfortunately, do not permit us to determine whether individual Ss did in fact respond this way. In any event, if we assume that \$1.00 is a plausible, but \$20.00 an implausible, reward, then the results fall neatly into the pattern of all previous and more extensive experiments on the effect of credulity on pressures to conformity (Fisher & Lubin, 1958).

To sum up, the design of this experiment does not allow us (a) to check whether discrepant cognitions were in fact produced, and (b) to rule out alternative explanations.

Incidentally, the authors of other related studies (Brehm, 1960; Brehm & Lipsher, 1959; Cohen, Terry, & Jones, 1959) have difficulty in accounting for all of their results according to dissonance theory predictions. These difficulties disappear if we use a plausibility explanation. The argument would proceed along these lines: If an individual is subjected to many pressures towards change from a number of different sources, (a) each pressure will act on the individual, and (b) their effect will be cumulative. For instance, we may increase pressure on an individual by limiting his freedom of choice, by giving him acceptable rewards, by presenting him with statements that strongly support a position discrepant to his own, by increasing the size of the discrepancy, and so on. Each of these alone will produce a greater and greater opinion change until—and this is the critical part of the argument—the situation becomes implausible, at which point the S will ignore the pressures and show no change. It also seems reasonable to suppose that a combination of *any* of these factors will act cumulatively to produce the implausible effect.

We can express this situation in statistical terminology. For example, if we have a two-factor experiment with two levels of pressure towards opinion change in each variable, we would expect to find that the two main effects are significant. Moreover, we would predict that the interaction would also be significant primarily because the combination of both "high pressures" would be implausible and so produce the least opinion change. In general, this is the pattern of results obtained by the dissonance workers in experiments of this type.

*Mealtime Troubles.* Another example of an untested interpretation occurs in the Brehm (1959) experiment on the effect of a *fait accompli* in which boys were offered a prize if they ate a portion of a disliked vegetable. While eating it, some Ss were casually told that a letter would be sent to their parents informing them of their participation in the experiment and of the vegetable they ate. Those boys who indicated they had trouble about eating the vegetable at home (i.e., it was more often served than eaten) subsequently changed their rating of the disliked vegetable towards a more favorable one. What did the letter mean to these boys? According to the author it meant that "the Ss would have to eat more" of the vegetable at home. But this is a guess, not based on any evidence in the experiment. Furthermore, in an extension of the same procedure at the same school with equivalent Ss from the same classes, direct manipulation of the commitment to further eating "failed to produce an overall effect on liking [Brehm, 1960, p. 382]." Under the circumstances, we find it difficult to accept the author's contention that the *fait accompli* increased cognitive dissonance by increasing the commitment to eating. There is little doubt that mentioning the letter

changed the ratings, but only for boys who had mealtime troubles. The key to the problem most likely lies in the expectation these boys had about the effect of the letter on their parents and on themselves. However, the design of the experiment does not allow us to find out what this expectation was.

*Confounded Effort.* In a recent experiment, Aronson (1961) tried to separate the effects of secondary reinforcement from dissonance in a rewarding situation.

Reinforcement theory suggests that stimuli associated with reward gain in attractiveness; dissonance theory suggests that stimuli associated with "no reward" gain in attractiveness . . . if a person has expended effort in an attempt to attain the reward [p. 375].

Aronson argues that since the effect of secondary reinforcement is constant, nonrewarded objects should become more attractive as the effort to obtain them increases.

In order to test this hypothesis, Ss fished for cans to obtain a reward (\$.25) inside one third of the cans. The rewarded cans were of one color, the non-rewarded ones of another, but the Ss could not determine which they had snared until the cans had actually been pulled out. One group of Ss—the low-effort group—was told that their task was not tedious. They had the relatively easy task of fishing out a can with a magnet, a task which took them, on the average, only 14 seconds per can. Another group of Ss—the high-effort group—was told that their task was extremely tedious. They had the relatively difficult task of fishing out a can with a hook, which took them, on the average, 52 seconds per can. All Ss continued fishing for the reward money until 16 unrewarded cans had been pulled out. The Ss rated the relative attractiveness of the two colors before and after carrying out the task. The results show that

in the low-effort condition, the attractiveness of the color on the rewarded cans increased (a secondary reinforcement effect), but in the high-effort condition no change was observed. All of this was interpreted as substantiating the cognitive dissonance predictions.

Aronson explains the lack of change in the high-effort condition by saying that the effects of dissonance and secondary reinforcement are equal but opposite in direction, and so cancel each other. However, if we look more carefully at the experimental manipulation of effort, we see that the low-effort condition is actually a reward rate of \$.25 about every 42 seconds, and the high-effort condition is actually a reward rate of \$.25 about every 156 seconds. In other words, the low-effort group is, at the same time, a high-reward-rate group, and the high-effort group, a low-reward-rate group. The difference obtained between the two groups could then be simply the result of the difference in reward rates, and the lack of change in the high-effort group, the result of their low reward rate.

To summarize, Aronson tried to demonstrate the effect of effort in a rewarding situation. However, the design of the experiment confounds effort with reward rate. As a result, no unambiguous conclusions can be drawn as to the effect of effort.

There is yet another experiment on effort in which confounding occurs. Yaryan and Festinger (1961) tried to show the effect of "preparatory effort" on belief in a future event. The Ss volunteered to participate in an experiment labeled "Techniques of Study" which was supposed to investigate the techniques, hunches, and hypotheses that students use to study for exams. The Ss were told that only half of them would take part in the complete experiment which involved taking an IQ test. All Ss

were given an information sheet on which there were definitions essential to this supposed IQ test. In the high-effort condition, Ss were told to study the sheet and memorize the definitions. In the low-effort condition, Ss were told to glance over the definitions briefly. The latter were also told that they would have access to the sheet later if they were to take the IQ test. Each S was then asked to express his estimate of the probability that he would take the IQ test. The results show that Ss in the high-effort group thought it was more probable that they would take the test.

The authors (Yaryan & Festinger, 1961) explain the results in the following way: Exerting "a great deal of effort" is inconsistent with the cognition that one may not take the test, so Ss in the high-effort group should believe "more strongly in the likelihood of the occurrence of the event [p. 606]." This might very well be the case, but in this experiment the variable of effort is confounded with the presence of other predictors for the event. All Ss had been told that this was an experiment on the techniques of study, but the only group which *did* any studying was the high-effort group. In addition, the studying that was done was highly relevant for the IQ test. Under the circumstances, it does not seem at all surprising that Ss in this group took these additional cues to mean that they were assigned to the complete experiment and to the IQ test. As it stands now, the Yaryan and Festinger experiment does not separate the effect of effort from that of additional cues.

*Reliability Is Not Validity.* As we have seen, most cognitive dissonance formulations are concerned with what happens after a person makes a decision. One of the earliest experiments designed specifically to investigate this problem was the gambling experiment

described by Festinger (1957, p. 164) and successfully replicated by Cohen, Brehm, and Latané (1959) with minor variations in procedure. The agreement between these two studies has done much to enhance the belief in the validity of cognitive dissonance formulations (e.g., Riecken, 1960, p. 489).

The experimental procedure in these two studies was relatively simple. Each S played a card game with the experimenter for variable money stakes. Before beginning the game each S was informed of the rules of the game, and on the basis of this information chose one of two sides on which to play. He was told that he could change sides once during the game, but that it would cost him money. He was led to expect that he would play 30 games. At the end of 12 games, play was interrupted and S was given a probability graph to study. The graph, a different one for each side, gave the (false) information that the chosen side was the losing one. The dependent variables were the time spent in studying the graph and the number of people who changed sides. Results were analyzed in terms of a weighted average of the amount of money won or lost. The pattern of results obtained is very complex and would require at least a fourth-order parabola to describe it. Nevertheless, the various ups-and-downs were interpreted as supporting the dissonance theory predictions for postdecision, information seeking processes.

Two things strike us about the dissonance theory interpretation of this experiment. First, Festinger is not consistent in his dissonance formulations. Let us look at the way in which the results are interpreted. The money winners spent a moderate amount of time studying the graph. Festinger considers this the result of dissonance produced by the information in the graph which purported to show that these winners

were actually on the losing side. If we accept this line of reasoning, it should follow that the losers would have no such dissonance (the graph confirmed their losses), and would therefore spend little time on the graph. This was not so. Festinger (1957), however, has three other dissonance explanations to account for the complex behavior of the losers. He argues, first of all, that loss of money is itself dissonance producing and the bigger the loss, the bigger the dissonance. The small losers spent as much time as they did in the "hope that the graph would tell them they were actually on the correct side [p. 171]." If this explanation is correct, the bigger losers should have spent an even longer time searching the graph—but they did not. Festinger explains this away by saying that the bigger losers would avoid the graph "if the graph were perceived as yielding information which would probably increase the dissonance which already existed [p. 172]." If this explanation is correct, the biggest losers should have spent the least amount of time on the graph—but they did not. To explain this behavior Festinger postulates yet another hypothesis. For the biggest losers "the easiest way to eliminate the dissonance would be to increase it temporarily to a point when it was greater than the resistance to change of the behavior," that is, they would study the graph, then switch sides. If this explanation is true, then we would expect that all of the biggest losers, and only the biggest losers, would switch sides—but this was not so. It should be noted that these four dissonance hypotheses are inconsistent with each other, since they predict effects in different directions. Moreover, there is no *a priori* way of determining the degree to which each particular hypothesis applies to the groups. This whole matter can be summarized in another way: If

the pattern of results had been exactly the reverse, these same explanations would apply just as well.

This brings us to the second point. The most important criticism of this gambling experiment is that it is not so much an experiment on the dissonance-reducing effects of information in post-decision processes, as it is an experiment on "information seeking in predecision processes" (suggested by F. E. Emery). The *Ss* had been told that they could change sides and they were actually given an opportunity to do so when they were handed the graph. Festinger (1957) and Cohen et al. (1959) reported that many *Ss*, both winners and losers, announced their decision to switch sides at this time. What was not reported, however, was the number of *Ss* who looked at the graph in order to reach a decision whether or not it would be more profitable to change sides. In other words, *Ss* looked at the graph not to reduce dissonance, but to look for information to help them decide whether they should change sides. With this interpretation, the pattern of results becomes more obvious and reasonable. For example, in one of the two conditions of the Cohen et al. replication those *Ss* who neither won nor lost any money spent the most time looking at the graph. This finding is entirely inexplicable in dissonance terms, even with an imaginative use of all four of Festinger's hypotheses. However, if we consider this experiment to be concerned with predecision processes, then we see that these *Ss* had gained the least amount of information from the actual play of the game and were, therefore, trying to extract as much information as possible from the graph before reaching a decision.

Taking all of these factors into consideration, we are forced to conclude that Festinger's interpretations, however in-

genious, are unnecessarily elaborate and unjustified. Moreover, the successful replication of the experiment suggests—not that the cognitive dissonance formulations are valid—but only that the results of experiments of this type are reproducible.

Subsequent experiments on selected aspects of information seeking in post-decision processes have done nothing to clarify the situation. Adams (1961) and Maccoby, Maccoby, Romney, and Adams (1961) showed that people tend to seek information which agrees with their viewpoint, but Rosen (1961) obtained results which show that people tend to seek information which disagrees with their viewpoint. It seems more likely to us that, in general, people will often seek new information, whether it be consonant or contrary. Indeed this is apparently the kind of result that Feather (1962) obtained. To be sure, when prejudice or some other highly motivated state is involved, people are selective in their perceptions and avoid contrary material. But under these conditions it is the motive itself, and not dissonance, that seems to be crucial.

#### *Interpretation of Manipulations*

Perhaps the illustrations cited above will suffice to show that the experiments adduced to support the theory of cognitive dissonance involve highly complex

manipulations. The effects of these manipulations are open to alternative explanations which have generally not been dealt with adequately by the authors. We can diagram this in the following way: Let us suppose that the complex experimental manipulations produce the Cognitions 1, 2, 3, 4 . . .  $n$  in the  $S$ , as illustrated on the left side of Figure 1. Two of these cognitions (Cognitions 1 and 2) are chosen by the experimenter as being the relevant discrepant cognitions producing dissonance. Any observed change in the dependent variable is then attributed to that dissonance. But, as we see by examining the whole of Figure 1, this may not necessarily be the case. Any one cognition, or any combination of cognitions, could have been responsible for the change in the dependent variable. There is no way of ascertaining which, because the effects of all these cognitions have been confounded.

It is possible to design experiments so that these effects are not confounded. As a step in this direction we recommend first of all that the experimental manipulations be simplified. It is difficult to agree about differences in cognitions when the instructions, task, and procedure differ in many ways for control and experimental groups. Our second recommendation is that additional control groups be included in the design

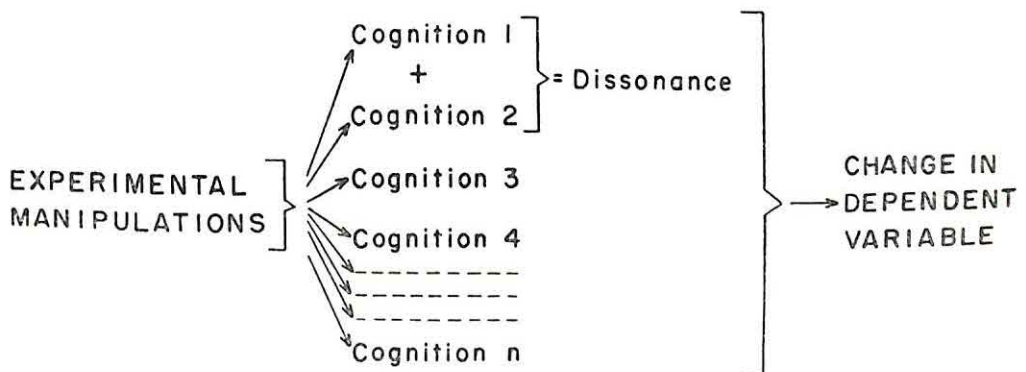


FIG. 1. The type of confounding frequently found in experiments on cognitive dissonance.

of these experiments to deal with the irreducible differences in experimental manipulation. Our third recommendation is that a little more attention be given to discovering the possible cognitions that an *S* might have about the situation, particularly those which might be contrary to dissonance theory. Only under such carefully controlled conditions can we begin to talk about unequivocal evidence in support of cognitive dissonance theory.

#### CONTROVERSIAL TREATMENTS OF THE DATA

So much for experimental manipulations. Now to see if the experimenter really got the results he says he did. Our most serious criticisms of the experiments cited in support of dissonance theory fall under the heading of methodological inadequacies in the analyses of results. Of these inadequacies the most important is the rejection of cases, not only because it is so fundamental a flaw, but also because the supporters of dissonance theory so often do it.

#### *Rejection of Cases*

If an experimenter is interested in the performance of only a certain group of *Ss*, it is legitimate for him to select these *Ss* before beginning the experiment, or sometimes even after the experiment, before the results are analyzed. However, when *Ss* are selectively discarded after the data of an experiment have been collected, tabulated, and sometimes even analyzed, it leaves the reader with a feeling of uneasiness. The uneasy feeling grows if the *Ss* are discarded because their results are said to be "unreliable," or if the experimenter gives inconsistent reasons—or no reasons at all—for their rejection. But let us look at the experiments themselves.

*Unreliable Ss.* Brehm and Cohen (1959b) asked sixth-grade children to

indicate how much they liked several different toys before and after they had chosen one for themselves. The choice of the gift and the postchoice rating were made a week after the prechoice rating. The authors hypothesized that there would be an increase in the evaluation of the chosen article, and a decrease in the evaluation of the nonchosen article, the greater the dissimilarity among the toys, and the greater the number of alternatives from which to choose. In general these predictions were upheld. But of the original sample of 203 children *only 72* were used in the analysis. In the authors' (Brehm & Cohen, 1959b) words, the reasons for the reduction were as follows:

First, the choice alternatives for each *S* had to be liked, but not so much that an increase in liking would be impossible. Second, one alternative had to be initially more liked than any other so that its choice could be expected. Increased ratings of the chosen item are thus not likely to be simply a result of normal (and random) changes in actual liking from the first questionnaire to the second. *In addition, Ss who failed to choose the alternative initially marked as most liked, were excluded because they gave unreliable or invalid ratings.* Finally, in order to ensure that initially less liked alternatives were seriously considered as possible, initial ratings of these alternatives could not be much lower than the most liked alternative [p. 375; italics added].

Note first that the exact limits of all these requirements were determined only after inspection of the data, despite the fact that each *S* had been given a prearranged choice based on his initial ratings. However, let us look more carefully at the italicized item—the exclusion of *Ss* because of unreliability. If *Ss* give unreliable results, it is usual to assume that the measuring instrument itself is unreliable; indeed, the authors themselves admit this when they mention "the low reliability of our measure of liking." However, discarding selected *Ss* does nothing to improve the reliability of an instrument.

Discarding Ss who did not choose the alternative initially marked as most liked may in fact falsely reduce the computed error variance, change the mean values, and so enhance the possibility of obtaining a significant difference in rating. To illustrate, the upper half of Figure 2 shows the ratings for two toys, X and Y, which satisfy the conditions specified by Brehm and Cohen: they are both liked, one is liked more than the other, but the difference is not great. Now let us assume that the ratings are subject to errors of measurement and that they vary randomly from time to time. Let us further assume that the expressed rating, the liking, at any one moment in time, is perfectly correlated with choice.

The situation a week later is shown in the bottom half of Figure 2. The ratings for X are now distributed as  $X'$ , the ratings for Y, as  $Y'$ . Now the Ss are asked to make a choice. Let us assume the null hypothesis, that is, the actual process of choosing a gift does not alter the liking or rating of a toy. Since the choice and the postchoice rating occur so close together in time we can also assume that no random change occurs from just before the choice to just after

it. If, as Brehm and Cohen did, we discard all those Ss who chose Y rather than X, this means that we eliminate from the shaded area in the bottom half of Figure 2 all those cases in which  $Y' > X'$ . Such a process can only reduce the variance of both distributions of differences in ratings (that is,  $X' - X$  and  $Y' - Y$ ), and automatically increase the mean difference between them. Moreover, this selection procedure will automatically produce precisely the effect which the authors predicted, namely, that the mean rating for the chosen toy increases, while the rating for the non-chosen toy decreases. Furthermore, these effects will be greatest in the group from which most such Ss were discarded. Most of these discards came from the "four alternatives condition,"<sup>2</sup> and the change in rating is actually greatest for this condition.

In a footnote (p. 376) the authors (Brehm & Cohen, 1959b) state they carried out similar computations on their "unselected population," that is, on the entire sample of 203 Ss. Although the sensitivity of the statistical test is now nearly twice as great (because of the increase in  $N$  from 72 to 203), they find

<sup>2</sup> J. Brehm, personal communication, 1961.

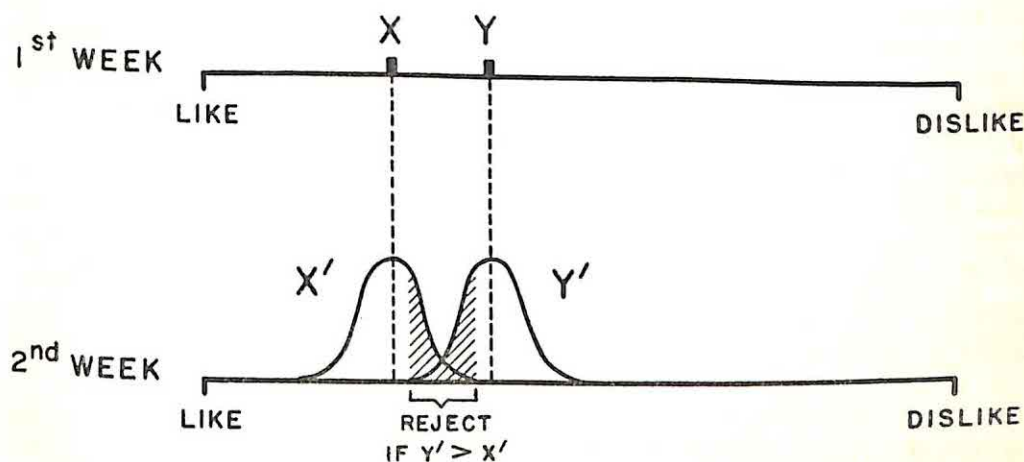


FIG. 2. This illustration shows how the rejection of Ss from the shaded area may have introduced a statistical artifact into the experiment by Brehm and Cohen (1959b).

no effect due to the number of alternatives (one of the two predictions made). Tests of the other prediction "yield support" for the "dissimilarity hypothesis." It is not clear, however, whether the authors mean by this a statistically significant difference, or simply a nonsignificant trend. To sum up, it seems reasonable to conclude that the significant results obtained in this and similar experiments may very well be statistical artifacts.

*Contradictory Reasons.* Sometimes it is difficult to reconcile the reasons given for the rejection of cases with other statements by the same author. For example, in the experiment on "Attitude Change and Justification for Compliance" (Cohen, Brehm, & Fleming, 1958), the initial analysis showed no significant difference between the two justification groups. The authors then eliminated more than half of the Ss (47 out of 92), carried out a second analysis on the remainder, and concluded that "the difference in amount of change is significant by one-tailed  $t$  test at .07 level." Relatively more Ss whose opinion did not change were eliminated from the low-justification condition (35 out of 63). Not surprisingly, the new mean for the low-justification condition turned out to be greater than for the high-justification condition.

Part of the reasoning for this selection of cases was as follows (Cohen et al., 1958): "since extremity of position inhibits attitude change . . . it seems reasonable to eliminate the extremes [p. 277]." A year later, however, Cohen (1959) made this statement: "If the individual . . . engages in some behavior with regard to the contrary communication . . . then the greater the discrepancy [extremity of position], the greater the opinion change [p. 387]."

In their original article, Cohen et al. state that their results should be in-

terpreted cautiously, but, unfortunately, they do not follow their own advice. Whenever these authors refer to their findings in later articles (e.g., Brehm, 1960; Cohen, 1960), they quote their results as substantiating cognitive dissonance theory without any of these cautionary reminders.

*What Is Going On?* An example of sample reduction for obscure reasons occurs in an experiment on the readership of "own car" and "other car" advertisements by new and old car owners (Ehrlich, Guttman, Schonbach, & Mills, 1957). A group of 65 new car owners was randomly chosen from a list of recent auto registrants. The car advertisements read by this group were compared with those read by a group of 60 old car owners chosen from a telephone directory. The raw data for these analyses were the percentages of car advertisements noticed and read in a selection of magazines and newspapers which the owners had previously indicated they read regularly. The cognitive dissonance theory predictions were that new car owners would most often read advertisements about their own make of car and avoid reading those of competing makes. In general, these predictions were upheld for the data presented.

The principal difficulty with this experiment is that cases were successively rejected in various stages of the analysis so that when one finally arrives at the critical statistical test it is virtually impossible to determine what the remaining data mean. Let us see if we can trace the authors' steps in this process. The authors first present a table showing the mean percentage of advertisements noticed and mean percentage of advertisements read of those noticed for each of the categories "own," "considered," and "other car." They (Ehrlich et al., 1957) state in a note accompanying the table that:

The  $N$ 's are reduced because in some cases no advertisements of a particular kind appeared in the issues shown or none of those which appeared were noticed. They are further reduced because not all respondents named cars as "seriously considered" [p. 99].

The first and third reasons impose a limit on the number of  $S$ s whose results could be used. The largest reduction due to these two limitations was in the category "considered car" for old car owners, where the  $N$  of 60 was reduced to 31.

The second reason given in the quotation above means that an owner who did not notice any advertisement in a particular category was discarded from the table of "advertisements read" for that category. For example, if an owner noticed (or noticed and read) an ad about another car but did not notice any advertisements about his own car, he was included under the category of "other car," but excluded from the category of "own car" in computing the mean percentage of "advertisements read of those noticed." Up to one third of the remaining cases were eliminated from the various categories for this reason.

The next point at which still more cases are rejected is in the computation of several sign tests of significance. We are told that the  $N$ 's are reduced because not all comparisons were possible. What this means is that significance tests were computed only on those owners who *noticed at least one advertisement in each of the pairs of categories compared*. Finally, those owners who read an equal percentage of advertisements in each of the two categories were also discarded.

Taking all of the above factors into account we find that as much as 82% of the original sample was discarded in certain categories!

The sign tests mentioned above were used only for making certain pairs of comparisons. For overall tests of their hypotheses the authors resorted to chi

square and give terminal chi square values, with their associated probabilities, alongside the tables for the sign tests. The article itself does not say upon what  $N$ s, or what groupings, the chi squares were computed, but correspondence<sup>3</sup> reveals that the chi square tests were made on the same  $S$ s as were used in computing the sign tests.

At best this entire situation may be described as unclear. In the first place, it is difficult to know how to interpret significance tests based on such highly selected data. Furthermore, in computing chi squares for the same  $S$ s as were used in the sign tests, it appears that the authors discarded some data (the ties) which should properly have been included. If we have been able to thread our way correctly through the authors' manipulations of the data we find that the chi squares, computed for all the relevant data, are less significant than reported by the authors, and, in two of three cases, change a nominally significant value to a nonsignificant one. In any event, there can be no doubt that the authors' (Ehrlich et al., 1957) summary statement "It was found that new car owners read advertisements of their own car more often than . . ." needs considerable qualification. With so much selection of  $S$ s and with such intricate manipulations of the data, some of it never fully explained, one can hardly describe the results as *public*, or the findings as necessarily significant.

*Manipulation Not Successful.* Still another type of rejection we find in these studies is the elimination of entire groups of  $S$ s. If one variant of the manipulation fails to show an effect, it is not legitimate to discard all the  $S$ s in that group from the analysis. The analysis should properly be carried out on all the data and the interpretations should be based on the complete analysis.

<sup>3</sup> J. Mills, personal communication, 1962.

sis. Brehm (1960), for example, used reports on the vitamin and mineral content of vegetables to try to influence the attitudes of Ss after they had eaten a disliked vegetable. One group of Ss received the vitamin report, the other group the mineral report. Since the mineral report "failed to affect" the dependent variable "the results for these subjects [were] omitted from this report [Brehm, 1960, footnote, p. 380]." One consequence of rejecting an entire group is that we do not know if there is a significant interaction between type of report and the other variables. Until this is established, it is misleading to consider a segment of the results as significant. In addition, the author nowhere states that his findings are specific to one type of report only. His summary is in terms of "communications about food value."

*Reallocation Instead of Rejection.* An interesting variant of the rejection of Ss occurs in the Raven and Fishbein (1961) study on the effect of "Acceptance of Punishment and Change in Belief." Groups of Ss were run under two conditions, "shock" and "no-shock." There were 13 females and 13 males in each of these two conditions. The results show that there was no overall difference between the shock and no-shock groups. However, when the results were tabulated separately for the two sexes, it appeared that the female Ss in the shock group changed in the predicted direction, but that the male Ss did not. Here is how the authors (Raven & Fishbein, 1961) dealt with the situation:

Overall analysis of variance and interaction was not significant. Assuming that male shock Ss were part of a common population with the non-shock subjects, with respect to dissonance, an analysis of variance was conducted which showed the female shock subjects to be significantly different from the others [p. 415].

In other words the authors disposed of

the Ss who did not conform to their prediction, not by rejecting them, but by reallocating them to another group, the no-shock group. If females in the shock group really are significantly different from all the others, this should show up in a significant interaction. It does not.

Rejection of cases is poor procedure, but reallocation of Ss from experimental to control group, across the independent variable, violates the whole concept of controlled experimentation.

*Danger of Rejecting Ss.* A theme common to many of these rejections is that the unselected sample "does not permit of an adequate test" of the dissonance hypothesis. We are told:

In a social influence situation there are a number of potential channels of dissonance reduction, such as changing one's own opinion, changing the opinion of the communicator, making the communicator noncomparable to oneself, seeking further support for one's position, dissociating the source from the content of the communication, and distorting the meaning of the communication [dissonance theory position ably summarized by Zimbardo, 1960, p. 86].

Such a theoretical formulation is indeed all-encompassing and it provides a rationale which certain other dissonance theory workers have used for rejecting cases. The reasoning goes like this: If some Ss do not follow the specific predictions in a particular experiment (for instance, if they fail to show any opinion change) then those Ss are probably reducing their dissonance through some other channel or else they had little dissonance to begin with. If either of these conditions holds it is legitimate to exclude these Ss from the analysis since they could not possibly be used to test the particular hypothesis in the experiment. An inspection of results is considered sufficient to determine whether Ss are, or are not, to be excluded. Unfortunately, this line of reasoning contains one fundamental flaw: *it does not*

allow the possibility that the null hypothesis may be correct. The experimenter, in effect, is asserting that his dissonance prediction is correct and that Ss who do not conform to the prediction should be excluded from the analysis. This is a foolproof method of guaranteeing positive results.

Some people may feel that no matter how questionable the selection procedure, it must still mean something if it leads to significant results. This point of view, however, cannot be reconciled with the following fact of life: it is always possible to obtain a significant difference between two columns of figures in a table of random numbers provided we use the appropriate scheme for rejecting certain of those numbers. For all we know, selecting Ss so as "to permit an adequate test of the hypothesis" may have had precisely this effect. A significance test on selected Ss may therefore be completely worthless.

We strongly recommend that Ss not be discarded from the sample *after* data collection and inspection of the results. Nor is it methodologically sound to reject Ss whose results do not conform to the prediction on the grounds that they have no dissonance, or that they must be reducing it some other way. If there are any theoretical grounds for suspecting that some Ss will not show the predicted dissonance-reduction effect, the characteristics of such Ss, or the conditions, should be specifiable in advance. It should then be possible to do an analysis on all Ss by dividing them into two groups, those predicted to show dissonance reduction, and those predicted not to show it. If such a thing as dissonance reduction exists, it is theoretically and practically important to know the precise conditions under which it does and does not occur.

A summary of experiments in which Ss are rejected is given in Table 1.

TABLE 1  
LIST OF EXPERIMENTS FROM WHICH Ss WERE DISCARDED AFTER DATA COLLECTION

Experiment	Total <i>N</i>	Discarded (%)	Reasons given
Brehm (1956)	225	35	To permit adequate test of hypothesis 1. Unreliable Ss 2. Conditions not fulfilled
Brehm (1960)	85 <sup>a</sup>	38 <sup>a</sup>	One manipulated condition not significant
Brehm & Cohen (1959b)	203	65	To permit adequate test of hypothesis 1. Ceiling effect for high scorers 2. Adequate separation of choice points for dissonance to occur 3. Unreliable Ss
Brehm & Lipsher (1959)	114	10-14	None
Cohen, Brehm, & Fleming (1958)	92	51	To permit adequate test of hypothesis 1. Extremity of attitude inhibits attitude change
Ehrlich et al. (1957)	125	17-82	1. Material missing 2. Advertisements not noticed 3. Not all comparisons possible 4. Ties
Mills (1958)	643	30	To permit adequate test of hypothesis 1. Ceiling effect for high scorers 2. Honest improvers have no dissonance

<sup>a</sup> Estimated.

### *Refusals*

The previous section has been concerned with sampling bias due to the deliberate rejection of cases by the experimenter. There is another type of sampling bias, equally important but much more subtle, that occurs when Ss reject themselves from the study by refusing to participate.

In a recent review of cognitive dissonance experiments, Cohen (1960) concluded with what he considered was a "depressing" and "Orwellian" statement:

It could be said that when the individual feels that he has most freedom of choice, when his volition and responsibility are most engaged, he is then most vulnerable to the effects of persuasive communications and to all sorts of controlled inducements from the world at large [p. 318].

This statement follows hard on the heels of "the more negative the person is toward a communication or communicator, the more he can be expected to change his attitudes in the direction of the communication or communicator." These are indeed sweeping generalizations, particularly since they are based on the results of experiments in which from 4% (Cohen, Terry, & Jones, 1959) to as many as 46% (Rabbie, Brehm, & Cohen, 1959) of the total number of Ss refused to participate. Moreover, there is evidence in these studies that the Ss who refused to participate were actually those who had both the greatest freedom of choice and the strongest (most negative) views on the attitude in question. What actually appears to have happened is that those Ss with the strongest (most negative) views were so invulnerable to the effects of persuasive communications that they exercised their freedom of choice by walking out on the experimenter or refusing to comply in other ways. To take the results

of the remaining more vulnerable Ss and extrapolate from them to the population in general seems unjustified.

### *Inadequate Design and Analysis*

It is rare to find in this area a study that has been adequately designed and analyzed. In fact, it is almost as though dissonance theorists have a bias against neat, factorial designs with adequate Ns, capable of thorough analysis either parametrically or nonparametrically. The majority of their experiments are some variant of the  $2 \times 2$  factorial with unequal, nonproportional, and generally small Ns in each cell. These restrictions make it impossible for the authors to carry out ordinary analyses of variance. Instead we find them making use of a hodgepodge of *t* tests and a statistic which they refer to as an "interaction *t*" (Walker & Lev, 1953, pp. 159-160).

Making a number of ordinary *t* tests on the same set of data, without a prior overall test of the null hypothesis, can be misleading. The principal difficulty is that in making such multiple comparisons the experimenter is allowing himself a number of opportunities to find an event (significance) which normally occurs infrequently. As a result, the usual *t* tables underestimate the true probabilities, that is, the probabilities obtained suggest a level of significance which is higher than warranted. Another way of saying it is that if, out of several subgroups, one finds one or two *t*'s significant, he is, in effect, capitalizing on chance (e.g., Sakoda, Cohen, & Beall, 1954). A further complication arises if the interaction is significant, since this introduces the usual difficulties about interpreting the main effects (e.g., Lindquist, 1953, p. 209). Some of the special statistical problems involved in the "postmortem" testing of comparisons were, of course, being discussed in the

psychological literature well before dissonance theory appeared on the scene (e.g., McHugh & Ellis, 1955); but for an excellent discussion of the basic issues involved in making multiple comparisons, see the article by Ryan (1959). None of these problems is ever faced squarely by the writers in this field. As a result, the authors sometimes reach conclusions that are not really warranted.

*Examples.* We can illustrate these remarks by referring to an analysis carried out by Brehm (1960) on two treatment variables, commitment and communication. There are three levels of commitment—control, low-eating, and high-eating—and, in addition, two types of communication—support and no support. Since the *N*s for these six groups are different (they vary from 7 to 11), it is not possible to carry out an ordinary analysis of variance. With such data at least 15 *t* tests and 3 interaction *t*'s are possible. Brehm gives the results of 7 such *t* tests (4 are nominally significant) and 2 such interaction *t*'s (both nominally significant). How do we interpret the results? Frankly, it is impossible. Taken at its face value, the analysis is not only useless, it is misleading.

An allied set of criticisms can be leveled at the analysis carried out by Brehm and Cohen in their study of the effects of choice and chance in cognitive dissonance (1959a). The design involved two types of relative deprivation, high and low. Five sections of an introductory psychology course were used as *S*s. The low- and high-deprivation conditions were experimentally manipulated and perceived as such by the *S*s. The low- and high-choice conditions were, however, determined separately for each section on the basis of their medians on the perceived-choice rating scale. Separate interaction *t*'s were calculated for each of the five sections. The *N*s in each

cell were very small, ranging between 3 and 10 with an average of about 7. The probability values for these 5 interaction *t*'s showed that one was significant, two tended to significance, and two were nonsignificant (one was actually a reversal). Here again the authors' failure to compute and report the results of an overall test make it exceedingly difficult for readers to interpret their findings. Moreover, there seems to be little justification for using a different value for the cutoff point between high- and low-choice for each section. In fact, such a procedure might in itself lead to statistically significant median differences between the sections. There may indeed be a significant interaction between choice and deprivation, but the evidence for it is, at best, questionable.

In dissonance experiments there is often a marked change between the pre- and posttest measures for both control and experimental groups. This is in itself an interesting phenomenon and should be thoroughly evaluated. An analysis should be complete—large main effects should not be ignored just because dissonance theory predicts only an interaction, or vice versa.

It is not impossible to apply a rigorous methodology to this area. Dissonance theorists would have done well to emulate the example set by Kelman as far back as 1953 (a study which, incidentally, anticipates and predates most of the areas of interest for cognitive dissonance workers). All of the problems that beset research in this area, such as unequal *N*s, class differences, and so on, were handled expertly by Kelman. More recently, such eclectic workers as Rosenbaum and Franc (1960) and McGuire (1960) have also been working in this area and have been using rigorous and comprehensive methods of analysis. In short, there appears to be no reason why

methodology in this area cannot be sharpened.

A summary of experiments in which the analyses and statistical interpretations are doubtful is given in Table 2.

### *Straining for Significance*

The final feature of the analyses that is apt to be misleading is the fact that authors tend to present results as significant and as supporting the dissonance theory prediction when the probabilities are greater than the usually accepted value of .05. Probability values between .06 and .15 (once even .50!) do not constitute striking support for any theory, particularly if it is preceded by a selection of Ss and poor analysis. It is also extremely disconcerting to find these statistically nonsignificant trends quoted authoritatively in subsequent reports and later reviews as substantiating the theory, without any qualifying statements.

### OVERALL EVALUATION

Having now reviewed much of the experimental work supporting cognitive dissonance theory, we conclude that, as a body of literature, it is downright disappointing. Too many studies have failed to stand up to close scrutiny. Yet it is also obvious that the dissonance framework has a seductive allure for many social scientists, an allure not possessed by the rather similar, but symbolically more complex, interpretations by Heider (1958), Osgood and Tannenbaum (1955), or Newcomb (1953).

### *Paradox of Simplicity*

The magical appeal of Festinger's theory arises from its extreme simplicity both in formulation and in application. But in our review we have seen that this simplicity was generally deceptive; in point of fact it often concealed

TABLE 2

SUMMARY OF SOME EXPERIMENTS WITH INADEQUATE DESIGN AND ANALYSIS

Study	Criticism of design and analysis
Allyn & Festinger (1961)	No control group (repeat attitude test, no talk); interaction significance not presented
Aronson & Mills (1959)	Overall significance not presented
Brehm (1956)	Maximum $N = 225$ , but regression equation based on $N = 557$ and $N = 534$
Brehm (1960)	Overall significance not presented
Brehm & Cohen (1959a)	Overall significance not presented
Cohen (1959)	No control group (repeat attitude test, no counterinformation); groups not equated on initial attitude
Cohen, Terry, & Jones (1959)	No control group (repeat attitude test, no new information); groups not equated on initial attitude
Ehrlich et al. (1957)	No control group (predecision car ad reading)
Festinger & Carlsmith (1959)	Overall significance not presented
Mills (1958)	Overall significance not presented
Mills, Aronson, & Robinson (1959)	No control group (preferences, but no decision); overall significance not presented
Rosen (1961)	No control group (preferences, but no decision)

a large number of confounded variables. Clearly much can be done to untangle this confounding of variables by careful experimental design. Nonetheless, there may still remain another problem more fundamental than this. In general, a cognitive dissonance interpretation of a social situation means that the relevant social factors can be condensed into two simple statements. To be sure, Festinger does not say formally that a dissonance theory interpretation works only for two discrepant statements; but it is precisely because in practice he does so limit it that the theory has had so much acceptance. Which brings us now to the crux of the matter: *is it really possible to reduce the essentials of a complex social situation to just two phrases?* Reluctantly we must say "No." To condense most complex social situations into two, and only two, simple dissonant statements represents so great a level of abstraction that the model no longer bears any reasonable resemblance to reality. Indeed the experimenter is left thereby with such emasculated predictors that he must perforce resort to a multiplicity of ad hoc hypotheses to account for unexpected findings. We see then that the most attractive feature of cognitive dissonance theory, its simplicity, is in actual fact a self-defeating limitation.

In conclusion, all of the considerations detailed above lead us to concur with Asch's (1958) evaluation of the evidence for cognitive dissonance theory, and return once more a verdict of NOT PROVEN.

## REFERENCES

- ADAMS, J. S. Reduction of cognitive dissonance by seeking consonant information. *J. abnorm. soc. Psychol.*, 1961, 62, 74-78.
- ALLYN, JANE, & FESTINGER, L. The effectiveness of unanticipated persuasive communications. *J. abnorm. soc. Psychol.*, 1961, 62, 35-40.
- ARONSON, E. The effect of effort on the attractiveness of rewarded and unrewarded stimuli. *J. abnorm. soc. Psychol.*, 1961, 63, 375-380.
- ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *J. abnorm. soc. Psychol.*, 1959, 59, 177-181.
- ASCH, S. E. Review of L. Festinger, *A theory of cognitive dissonance*. *Contemp. Psychol.*, 1958, 3, 194-195.
- BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, 52, 384-389.
- BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, 58, 379-382.
- BREHM, J. W. Attitudinal consequences of commitment to unpleasant behavior. *J. abnorm. soc. Psychol.*, 1960, 60, 379-383.
- BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, 58, 383-387. (a)
- BREHM, J. W., & COHEN, A. R. Re-evaluation of choice alternatives as a function of their number and qualitative similarity. *J. abnorm. soc. Psychol.*, 1959, 58, 373-378. (b)
- BREHM, J. W., & LIPSHER, D. Communicator-communicatee discrepancy and perceived communicator trustworthiness. *J. Pers.*, 1959, 27, 352-361.
- BRUNER, J. Discussion of Leon Festinger: The relation between behavior and cognition. In J. S. Bruner, E. Brunswik, L. Festinger, F. Heider, K. F. Muenzinger, C. E. Osgood, & D. Rapaport, *Contemporary approaches to cognition: A symposium held at the University of Colorado*. Cambridge: Harvard Univ. Press, 1957. Pp. 151-156.
- COHEN, A. R. Communication discrepancy and attitude change: A dissonance theory approach. *J. Pers.*, 1959, 27, 386-396.
- COHEN, A. R. Attitudinal consequences of induced discrepancies between cognitions and behavior. *Publ. Opin. Quart.*, 1960, 24, 297-318.
- COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, 56, 276-278.
- COHEN, A. R., BREHM, J. W., & LATANÉ, B. Choice of strategy and voluntary exposure to information under public and private conditions. *J. Pers.*, 1959, 27, 63-73.
- COHEN, A. R., TERRY, H. I., & JONES, C. B. Attitudinal effects of choice in exposure to counter-propaganda. *J. abnorm. soc. Psychol.*, 1959, 58, 388-391.
- EDWARDS, A. L. A critique of "neutral" items in attitude scales constructed by the method



- of equal appearing intervals. *Psychol. Rev.*, 1946, 53, 159-169.
- EHRICH, D., GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision exposure to relevant information. *J. abnorm. soc. Psychol.*, 1957, 54, 98-102.
- FEATHER, N. T. Cigarette smoking and lung cancer: A study of cognitive dissonance. *Aust. J. Psychol.*, 1962, 14, 55-64.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- FESTINGER, L. The psychological effects of insufficient rewards. *Amer. Psychologist*, 1961, 16, 1-11.
- FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, 58, 203-210.
- FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, 56, 230-238.
- HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, 21, 107-112.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, 6, 185-214.
- KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MACCOBY, ELEANOR E., MACCOBY, N., ROMNEY, A. K., & ADAMS, J. S. Social reinforcement in attitude change. *J. abnorm. soc. Psychol.*, 1961, 63, 109-115.
- MCGUIRE, W. J. Cognitive consistency and attitude change. *J. abnorm. soc. Psychol.*, 1960, 60, 345-353.
- MCHUGH, R. B., & ELLIS, D. S. The "post-mortem" testing of experimental comparisons. *Psychol. Bull.*, 1955, 52, 425-428.
- MILLS, J. Changes in moral attitudes following temptation. *J. Pers.*, 1958, 26, 517-531.
- MILLS, J., ARONSON, E., & ROBINSON, H. Selectivity in exposure to information. *J. abnorm. soc. Psychol.*, 1959, 59, 250-253.
- NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, 60, 393-404.
- OSGOOD, C. E. Cognitive dynamics in the conduct of human affairs. *Publ. Opin. Quart.*, 1960, 24, 341-365.
- OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychol. Rev.*, 1955, 62, 42-55.
- RABBIE, J. M., BREHM, J. W., & COHEN, A. R. Verbalization and reactions to cognitive dissonance. *J. Pers.*, 1959, 27, 407-417.
- RAVEN, B. H., & FISHBEIN, M. Acceptance of punishment and change in belief. *J. abnorm. soc. Psychol.*, 1961, 63, 411-416.
- RIECKEN, H. W. Social psychology. *Annu. Rev. Psychol.*, 1960, 11, 479-510.
- ROSEN, S. Postdecision affinity for incompatible information. *J. abnorm. soc. Psychol.*, 1961, 63, 188-190.
- ROSENBAUM, M. E., & FRANC, D. E. Opinion change as a function of external commitment and amount of discrepancy from the opinion of another. *J. abnorm. soc. Psychol.*, 1960, 61, 15-20.
- ROSENBERG, M. J. An analysis of affective-cognitive consistency. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm, *Attitude organization and change*. New Haven: Yale Univ. Press, 1960. Pp. 15-64.
- RYAN, T. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, 56, 26-47.
- SAKODA, J. M., COHEN, B. H., & BEALL, G. Test of significance for a series of statistical tests. *Psychol. Bull.*, 1954, 51, 172-175.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
- YARYAN, RUBY B., & FESTINGER, L. Preparatory action and belief in the probable occurrence of future events. *J. abnorm. soc. Psychol.*, 1961, 63, 603-606.
- ZAJONC, R. B. The concepts of balance, congruity, and dissonance. *Publ. Opin. Quart.*, 1960, 24, 280-296.
- ZIMBARDO, P. G. Involvement and communication discrepancy as determinants of opinion conformity. *J. abnorm. soc. Psychol.*, 1960, 60, 86-94.

(Received September 4, 1962)

# ELECTRICAL SELF-STIMULATION AND ITS THEORETICAL IMPLICATIONS<sup>1</sup>

C. R. GALLISTEL<sup>2</sup>

*Stanford University*

Evidence from electrical self-stimulation experiments is reviewed in the light of Deutsch's structural theory of behavior. The theory accounts for all salient observations; viz., high drive, little or no satiation, rapid extinction, the need for "priming" at the start of a day's trials, the difficulty in creating secondary reinforcers by using electrical stimulation of the brain as a primary reinforcer, the fact that different electrode placements interact with different natural drives, and the separability of the motivation and reinforcement processes. Other theories are briefly reviewed.

In 1954 Olds and Milner discovered that electrical stimulation of certain areas in the brain was rewarding; i.e., animals would learn to seek out such stimulation. If, for example, rats could stimulate certain parts of the septum, the hypothalamus, and the tegmentum by pressing a lever in a Skinner box, they would learn to press the lever at quite a high rate for long periods of time.

Many of the results of the research prompted by this discovery are confusing and somewhat paradoxical, as Zeigler (1957) noted when he first reviewed the literature. It has been found that although animals show a strong drive for electrical stimulation during experimental sessions, responses learned for this stimulation will extinguish with extreme rapidity. Likewise, the performance of animals in a maze where they are rewarded with brain shock will improve steadily during single experimental sessions with massed trials, yet

the performance declines markedly between the last trial in one day's session and the first trial of the next day's session. Finally, in spite of the evidence for strong drive and for high reward from electrical stimulation of the brain, most experimenters have apparently found it difficult to maintain responding on VI schedules (variable intervals of time between reinforcements) with a mean higher than 16-30 seconds, or on FR schedules (fixed number of responses between reinforcements) higher than 10:1.

Many theorists have cited the phenomenon of self-stimulation as support for hedonic theories of motivation and reward, but until such theories are elaborated it will remain hard to judge whether the data do in fact support them. That these areas are in some way involved in reward (or better reinforcement) is obvious from the very fact that they produce learning. To say that the sensations produced by electrical stimulation in these areas are pleasurable does not explain how these sensations operate to produce learned behavior. Nor does it by itself explain the paradoxical results noted above.

In order for the results of electrical stimulation studies to shed much light on the mechanisms of drive and reward, it is necessary to have some clearly

<sup>1</sup> This work was wholly supported by United States Public Health Service Grant M-4563 and National Science Foundation Grant G-21376.

I wish to thank J. A. Deutsch and C. I. Howarth for their many helpful suggestions and comments. My thanks also go to Patricia Shanahan for her invaluable help in finding bibliographical material.

<sup>2</sup> Now at Yale University.

worked out theories about the nature of these mechanisms. Such theories must provide answers to two questions: (a) How does the electric shock act on the central nervous system to organize the animal's behavior in such a way that it can reproduce the actions which led to the shock? (b) Whence comes the motivation which activates the response patterns established by the shock?

This paper will evaluate the results of electrical self-stimulation in the light of a structural theory of behavior proposed by Deutsch (1960). The theory is quite rigorous, and the data appear to accord well with its predictions. Only that part of the theory is given which is directly relevant to an explanation of the characteristics peculiar to the phenomenon of self-stimulation.

The theory postulates that there are centers in the brain sensitive to bodily deficits or other motivating conditions, which centers normally generate motivational excitation. This excitation flows from these centers along pathways to the sensory-motor units of the brain, which determine the behavior pattern that this excitation will energize.

The motivational centers have sensory inputs which signal the arrival of goal objects such as food, water, sexual gratification, etc. These inputs have two effects: First, they gradually inhibit the sensitivity of the motivational centers, rendering them less sensitive to the conditions which excite them. This accounts for the cessation of consummatory activities. Second, when the center receives the goal signals, it sends out reinforcing impulses that hook up the neural units which mediated the actions that immediately preceded the goal signal. Thenceforth, motivational excitation from the primary center flows to these units via the connections established between them by the reinforcing impulses. Near the primary center the

reinforcing impulses travel on pathways which run alongside the motivational pathways from that particular center.

The theory assumes that electrode placements which produce self-stimulation are located in these motivation and reinforcement channels. A shock from such an electrode has two effects. It introduces artificial excitation into *both* the reinforcement and the motivation pathways. The impulses in the reinforcement pathways hook up the neural units which mediated the behavior leading to the shock. The impulses in the motivation pathways excite these units so that the animal repeats the performance.

#### STRENGTH AND PERSISTENCE OF DRIVE

It should be emphasized at the outset of the review that although in some areas the results consistently point in one direction, in many other areas the results obtained depend for the most part on the placement of the electrode and the intensity of the stimulating current. One of the advantages of the structural theory is that it seems to account for most, if not all, of this variability in the results.

Following the structural theory, one would expect to see little or no satiation during continual electrical self-stimulation, since the reinforcing impulses do not themselves produce satiation. It is the incoming sensory impulses which inhibit the primary centers, thereby causing satiation. In general, the complete absence of satiation has been one of the most striking and consistent results of self-stimulation experiments. Olds (1958c) has found that rats will stimulate themselves continually for 24-48 hours until they drop from exhaustion.

There is only one report of satiation effects in the literature. Olds (1958d) reported that there is a significant de-

cline in the rate of lever pressing in rats with telencephalic placements after they have continually stimulated themselves at a fairly high rate for 8 hours. However, considering the rate at which they self-stimulate, 8 hours is an exceedingly long period. This may mean that the decline in rate is due to factors other than satiation.

The fact that animals will self-stimulate till they drop from exhaustion indicates that the drive for this stimulation must be fairly strong. This has been borne out in experiments by Olds (1958c) where he demonstrated that rats would cross a much more highly charged grid for electrical stimulation of the brain than they would for food when they were 24 hours hungry.

#### EXTINCTION

In view of the evident strength of the drive, it is puzzling that extinction should come so quickly. Yet, a close look at the data from most experiments shows that a very sharp extinction curve is the rule. Seward, Uyeda, and Olds (1959) reported:

Our results confirm earlier reports that this form of reinforcement produces steep extinction curves. It would be of interest to find comparable data for a drive-maintained response. *Unfortunately we have discovered no such data* [p. 298; italics added].

These results are not so puzzling in the light of the structural theory. Since the motivation for the responses is artificially injected by the shock itself, one might expect the responses to disappear soon after cessation of the shock, because the motivation for them disappeared and not because they were extinguished, i.e., unlearned. This also accords with the common observation that after "extinction" the animals can be recalled to the lever by giving them a few shocks (Olds & Milner, 1954).

In one case Olds (1956) has obtained something like a normal extinction curve.

However, these animals were running under 24-hour food deprivation. The motivation in this case may have come partly from the hunger center. The interaction between artificial motivation induced by brain shock and natural motivation will be discussed later.

Relevant here are data obtained by Seward, Uyeda, and Olds (1960) concerning massed and spaced trials. They found that performance during sessions of massed trials in a runway (with a lever in the goal box) improved markedly from trial to trial. There was, however, no trial-to-trial improvement in performance when the trials were spaced 15 minutes apart. The authors admitted that this is unusual: "Most studies on runway performance to food or water have reported no appreciable effect on acquisition due to spacing within the limits here used [p. 227]." They also found that the subjects extinguished quickly whether the lever was left in the goal box or withdrawn from it at the beginning of extinction. Finally, the rats given massed trials extinguished *more slowly* than the rats given spaced trials. Similarly, Olds (1956) found large decrements in performance in mazes from the last trial on one day to the first trial on the next. Nevertheless, taking each session as a whole, the animals showed day-to-day improvement. Here again, we might expect such results if we assume that decay of motivation over time and not reward is the key variable in spaced trials and between sessions.

In another experiment Olds (1958) found that his rats showed marked improvement in the first trial running speed over the first 4 days. However, his rats were 18-24 hours hungry. In this case hunger may have supplied the motivation. Nonetheless, he noted that the first trial of each day was *slower* than the last trial of the preceding day;

whereas for the controls who were running for food, the first trial was consistently *faster* than the last trial of the preceding day. The structural theory would predict that if the rats were run satiated, the day-to-day improvement in first-trial running speed would be much less marked and the decrement between the last trial of one day and the first trial on the next would be still larger.

In support of this hypothesis one may note that experimenters have apparently found it very hard to maintain responding on anything but small fixed-interval or fixed-ratio schedules. Sidman, Brady, Conrad, and Schulman (1955) reported schedules of VI=16 and FR=7:1. Brady and Conrad (1960), using a very high current, were able to keep their rats responding on a VI schedule with a 60-second mean. This compares poorly with VI schedules having means of 5–10 minutes or more, which will maintain slow but steady responding when the reward is food or water.

Monkeys seem able to maintain somewhat higher schedules. Brady (1960) maintained responding in monkeys at FR=50:1. Brodie, Moreno, Malis, and Boren (1960), again using a very high current (3 milliamperes), found one monkey which would respond on FR=150:1. This is the largest FR in the literature. Four of his eight monkeys, however, would not respond on a fixed ratio higher than 20:1. No one has shown that a partial reinforcement schedule increases the number of trials to extinction. The structural theory would predict that partial reinforcement would decrease the number of trials to extinction instead of increasing them, as it does when natural reinforcers are used.

An experiment by Howarth and Deutsch (1962) gives very strong support to the hypothesis that extinction of responses induced by electrical stimula-

tion is a function of drive decay and not of unlearning. They withdrew the lever from a Skinner box for periods of 2.5, 5.0, 7.5, and 10.0 seconds at the start of extinction. They then replaced the lever and allowed the rats to press it to some criterion of extinction. They found that the rats gave the same number of presses to extinction that they would have given, had they been permitted to press the lever (for nonreinforcement) during the period when it was removed from the box. For example, the rats pressed the lever an average of 1.92 times before reaching extinction after 7 seconds of nonreinforced lever pressing, and they pressed it an average of 1.87 times before reaching extinction after the lever had been removed from the box for 7 seconds at the start of extinction. The number of responses to extinction was purely a function of the time elapsed since the final brain shock and not of the number of unreinforced presses.

#### SECONDARY REINFORCEMENT

From the structural theory one should normally not be able to demonstrate the creation of secondary reinforcing cues by the use of intracranial stimulation as a primary reward. In the structural theory, shock produces the motivation. Hence, no shock, no response; unless some natural motivation is also present. This is in contrast to what one might expect from some sort of pleasure or incentive theory. The incentive theories might account for quick extinction, etc., by pointing out that the incentive was absent. One ought to be able to assume, though, that cue stimuli which had been associated with the incentive would serve to produce the behavior.

The evidence here is inconclusive. In general, though, it has been very difficult to show secondary reinforcement phenomena. Seward, Uyeda, and Olds

(1959) paired a light with all lever presses that produced shock, but found that the light did not increase the number of trials to extinction once the shock had stopped. This was probably not due to failure to discriminate or take note of the light, since it has been shown that rats are capable of discriminating tones concurrently with rewarding intracranial stimulation (Beer & Valenstein, 1960).

Stein (1958) did obtain some evidence of secondary reinforcement. He placed the rats in a Skinner box with two levers, one of which produced a tone and one of which did not. He recorded their operant rate of pressing each lever for 1 hour on 6 successive days. Then both levers were removed and the rats were given 100 brain shocks paired with the tone. The levers were again presented and the rate of pressing on each was again recorded. The brain-shock-positive group showed a significant preference for the tone linked lever and a significant increase in their absolute rate of pressing the tone lever. A brain-shock-neutral group showed little or no change in either preference or rate.

It should be noted, however, that the increased rate of pressing the tone lever in the experimental group is insignificant when compared with their rate of pressing for electrical stimulation itself. In a posttest to distinguish between the control and the experimental rats (the brain-shock-positive ones), the latter pressed a lever for electrical stimulation at a minimum of 9 times a minute or 540 times an hour. This compares poorly with the rate of 35 responses an hour obtained for the tone alone.

Since histological verifications and the details of pre-experimental conditions such as feeding are not given, one cannot rule out the presence of some natural motivation which the structural theory must assume. The theory would predict that the outcome of a secondary rein-

forcement experiment such as this one would vary with the manipulation of natural drives.

#### RELATIONSHIP BETWEEN EXPERIMENTALLY INDUCED AND NORMAL MOTIVATION

Mapping studies (Olds, 1960b; Olds, Travis, & Schwing, 1960) have shown electrical self-stimulation phenomena from a number of placements in the septum, the telencephalon, the hypothalamus, and the tegmentum. Olds thought he found two systems where electrical self-stimulation phenomena are present. The main system—the one which produces the highest rates of self-stimulation—is in the medial forebrain bundle and includes the medial anterior and posterior hypothalamus and the ventromedial regions of the telencephalon and the tegmentum. The second system, which produces lower rates, winds from the caudate and septal areas through the dorsal thalamus and into the tectum.

In the main or primary area, rate was a positive linear function of the stimulus intensity. In the secondary area rate was a square or fluctuating function of the stimulus intensity; i.e., as the stimulus intensity was increased the rate either went up and down or it quickly leveled off at a maximum. Olds theorized that placements producing positive linear correlation between rate and intensity were surrounded by reward cells for some distance so that the current spread at higher stimulus intensities simply affected more and more reward cells; that placements producing square functions were surrounded by a thin layer of reward cells with only neutral cells beyond these; that placements producing fluctuating functions were surrounded by concentric layers consisting alternately of reward and avoidance cells. Reynolds (1958) in a

similar study\* found only one pattern, namely, a steady increase in rate up to an "optimum" voltage followed by a steady decrease.

Hodos and Valenstein (1960) have pointed out, however, that the second system may produce lower rates simply because shocks in these areas also cause interfering motor effects such as twitches and crouching. They emphasize that the rate of pressing is not a good measure of the size or intensity of the reward. They gave their rats a two-lever choice situation in which they could obtain moderately strong stimulation from a septal electrode or low stimulation from a hypothalamic electrode. Their rats preferred the lever producing the stronger septal stimulation although the rate of pressing for the septal stimulation was slower than that for the hypothalamic stimulation.

Since the motivational channels in the brain must spread out from the motivational centers thought to be located in and around the hypothalamus, the results of the mapping experiments accord with what one might expect on the basis of the structural theory. One would also expect from the structural theory that the results obtained from different placements would be differentially affected by natural motivational states. The structural theory assumes separate reward and motivation channels for hunger, thirst, sex, fear, etc. Therefore, the performances for electrical stimulation obtained from one placement should be influenced by one of these naturally motivating states but not by the others, unless the electrodes were so large as to affect channels from two or more centers.

A number of studies give evidence that this is in fact the case. Brady, Boren, Conrad, and Sidman (1957) found that with certain placements, the rate of responding for electrical stimulation was considerably higher when the animals were 18-24 hours hungry and

thirsty. Unfortunately they did not separate the conditions of hunger and thirst. Hoebel and Teitelbaum (1962) implanted self-stimulation electrodes in the lateral hypothalamus. They found that ablating or anesthetizing the medial hypothalamus increased both eating and rates of electrical self-stimulation. Electrically stimulating the medial hypothalamus decreased both food consumption and rates of self-stimulation. When they applied the techniques of anesthetization, ablation, and electrical stimulation to certain parts of the *lateral* hypothalamus, they obtained the exact converse of the above results.

Olds (1958b), in an experiment on 16 castrated rats, found that both hunger and androgen levels influenced rates of self-stimulation. Further, he found that if hunger increased the rate of self-stimulation from one placement, then high androgen levels decreased it, and vice versa. This is what we might expect, since a high nonrelevant drive would interfere with a placement which was inserting artificial motivation into another drive system.

Deutsch and Howarth (1962) have demonstrated an interesting relationship between fear and self-stimulation from tegmental placements. They trained 10 rats to press a lever for electrical stimulation and then "extinguished" the response by holding the rats away from the lever for a minute or two. After a 5-minute pause during which the rats generally did nothing, they frightened them with a loud buzzer or a mild shock to the feet. The five tegmental animals immediately returned to the lever and began pressing it. They sometimes kept on pressing for several minutes. The other five rats did not return to the lever. This indicates that the tegmentum contains pathways for the motivational excitation arising from fear, and that a habit learned for electrical stimulation of the brain (ESB) can be reactivated by naturally arising motivation.

It is interesting to note in this connection that Olds (1958c) reported an avoidance area just beneath the tegmentum (see also Delgado, Roberts, & Miller, 1954). This may be a fear center. The hypothesis that the tegmentum contains fear pathways originating in or near this area may shed light on a puzzling result that Olds and Milner (1954) reported in their first paper. They found that one of their tegmental placements had an exceedingly high rate of operant responding even when it was not being reinforced by ESB:

Finally there is an electrode . . . between the red nucleus and the posterior commissure. It produced an acquisition score of 77%, but an extinction score of 81%. This must be a rewarding placement, but the high extinction score makes it difficult to interpret [p. 23].

It may well have been that the animal was quite frightened, in which case this dovetails nicely with the results obtained by Deutsch and Howarth.

Finally, Brady and Conrad (1960) ran an experiment testing for anxiety effects during self-stimulation. They sounded a tone for 5 seconds at the end of which they delivered a painful shock to the animal's feet. The tone thus becomes a conditioned stimulus (CS). Normally, sounding the CS will abruptly halt responding in a Skinner box until after the unconditioned stimulus (UCS). Their results, however, were very inconsistent. They ran the experiment on six rats, two cats, and four monkeys. They found no anxiety effect in the rats, but both cats demonstrated it. Some monkeys showed the anxiety affect and some did not. In one monkey they found no attenuation of responding from one placement (medial forebrain bundle), but marked attenuation from another (anterior thalamus). It is tempting to speculate that the placements where no attenuation was found were in fear pathways. Be that as it may, the authors (Brady & Conrad, 1960) concluded that their findings

would seem to provide strong support for the conclusion that the reinforcing properties of electrical stimulation of different parts of the brain may derive from quite different behavioral and physiological effects [p. 35].

The data, then, clearly support the hypothesis that there are channels of reward and motivation specific to the different natural drive states. But no one has yet mapped out what areas are related to which drives.

#### REWARD AS A SEPARATE PROCESS

There is good evidence to show that electrical stimulation of certain areas is motivating as well as rewarding. But what proof is there that reward and motivation are separate processes?

There is little direct experimental evidence on this score. Nevertheless, consideration of the indirect evidence makes this the most plausible conclusion. Thirst, hunger, sex, and fear motivation have been artificially introduced by electrical and chemical stimulation without any observed concomitant rewarding effects (Delgado & Anand, 1953; Larsson, 1954; Miller, 1957; Smith, 1956). In some cases this is probably due to failure to look for or test the rewarding properties. Margules and Olds (1962), using rats, found that all placements which produced increased eating also produced high rates of self-stimulation. It is clear, however, that in the case of sex and fear, motivation can be artificially produced by electrical stimulation without any concomitant rewarding effects (Miller, 1961). Hunger and thirst can also be produced by single injections of NaCl (Hoebel & Teitelbaum, 1962; Miller, 1961). The motivating effect of these injections is marked and fairly long-lasting. Their rewarding effect, if any, has never been observed. Since some electrode placements clearly produce both motivation and reward, it seems plausible to argue that they are two different processes. In the face of the evidence it would seem highly unlikely that the rewarding

effects of electrical stimulation could be accounted for by a drive *reduction* hypothesis. On the contrary, electrical stimulation from self-stimulation placements appears to increase, not to reduce, the drive for more.

The only direct experimental evidence comes from an as yet unpublished experiment by Deutsch, Howarth, Ball, and Deutsch (1962). They reasoned that if electrical self-stimulation placements are activating two separate processes or systems—a motivation system and a reward system—then the thresholds of electrical activation for the two systems might be different. A preliminary experiment indicated that the reward threshold was higher than the motivation threshold. It indicated that a current intensity just below the general threshold for electrical self-stimulation phenomena might be above the motivation threshold but below the reinforcement threshold. Consequently, animals stimulated by this current would show motivation by persisting in an already learned response but would not show any new learning.

To test this they compared performance under four different conditions of stimulation after an initial period of training on a lever pressing habit for electrical stimulation. Two conditions involved subthreshold stimulation; i.e., stimulation at intensities lower than those which were necessary to induce learning. In the first of these the stimulation was contingent upon a lever press—the rat received stimulation each time it pressed the lever and not otherwise. In the second condition stimulation was presented regularly by a pulse generator at a rate approximately equal to the rat's normal rate of responding for suprathreshold stimulation. Thus the animal received regular stimulation whether it pressed the lever or not.

The prediction here was that when the rats were receiving regular but not necessarily contingent stimulation, they

would respond more frequently than when stimulation was contingent upon lever pressing. Since electrically induced motivational excitation decays rapidly, any slight pause in responding would permit it to drop nearly to zero when stimulation is contingent upon lever pressing. Once motivation had dropped to zero, there would be little chance of restoring it because further stimulation is contingent upon pressing the lever and the rat would no longer be motivated to do so. However, if motivation is regularly replenished regardless of whether the animal is responding or not, small pauses in responding will not result in complete cessation of motivation. Seven out of eight rats did in fact respond more frequently while receiving regular *noncontingent* subthreshold stimulation than while receiving *contingent* subthreshold stimulation.

The second two conditions were the same as the first two except that stimulation was at threshold intensity. In this case the converse of the above results was expected. When threshold stimulation is given regularly by a pulse generator, it will be to a greater or lesser extent noncontingent upon the lever press; hence it will tend to reinforce either competing or at least interfering responses or "superstitions." On the other hand, contingent stimulation reinforces only the correct response, and, at threshold, it is strong enough to motivate a rapid rate of responding with little likelihood of complete decay of motivation during short pauses. Consequently, at threshold intensities responding should be more rapid with contingent stimulation than with noncontingent stimulation. This proved to be the case for all rats tested.

#### REWARD-AVOIDANCE PLACEMENTS

Several experiments have shown that animals find long shocks from some placements unpleasant, even though short shocks from these placements are

rewarding. It is difficult to say just what theoretical significance, if any, these results may have.

Roberts (1958) first reported the phenomenon. He found that he could train cats to learn the correct turn in a T maze in order to escape a continuous brain shock. When, however, he put the cats in an escape box and opened the door 15 seconds before giving them a continuous shock, they would not learn to leave the box in order to avoid the shock, although they did learn to escape from the box in order to terminate the shock. Roberts hypothesized that the onset of the shock was rewarding. To test this, he put the animals in a self-stimulation Skinner box. They did indeed learn to press the lever.

Miller (1957a) reported an experiment by Kirschner in which he taught rats to press one lever to produce a brain shock and another lever to terminate the shock.

The most complete study of the phenomenon was done by Bower and Miller (1958). They clearly confirmed initial hypotheses that continuous shocks from some placements *gradually* become unpleasant even though they are at first rewarding. As a start, they replicated Roberts' experiments and further demonstrated that rats *can* learn to leave a box in order to avoid an unpleasant brain shock provided that the placement is not rewarding for short shocks. They provided additional evidence by dividing their rats into two groups: those which showed reward-avoidance effects and those which showed only reward effects. They found that the reward-avoidance placements had a shorter mean duration of presses than did the pure reward placements when the brain shock continued as long as the lever was kept down. They also tested the rats under conditions where each bar press produced a 3-second shock. They found that the rate of pressing for the reward-avoidance placements decreased sharply,

whereas the pure reward placements showed no decrease in their rate of pressing. Finally, they did careful histological work which showed that the reward-avoidance placements were in the middle to anterior portions of the medial forebrain bundle and that the fornix was not involved. Placements in the far posterior parts of the medial forebrain bundle and other parts of the hypothalamus were purely rewarding. Olds and Peretz (1960) found that most of their reward-avoidance placements were located just below the medial lemniscus. Olds (1960a) reports reward-avoidance placements in the medial hypothalamus and the lower tectum.

Why placements in the forward part of the medial forebrain bundle produce unpleasant effects during long shocks is not known. The shock may spread to avoidance centers or it may spread to afferent pain fibers. Bower and Miller (1958) reported that some of their reward-avoidance placements showed clear signs of pain when they received long shocks. It may be also that the shock slowly affects the pain receptors in the blood vessels of the brain (if such receptors do exist).

#### DRUG STUDIES AND MISCELLANEA

Several experimenters have studied the effects of drugs on self-stimulation behavior (Miller, 1957b; Olds, 1958, 1961; Olds, Killam, & Eiduson, 1957). Since the effects of the drugs are themselves under investigation and still poorly understood, the results of these studies are more relevant to the literature on psychotropic drugs than to the literature which is primarily concerned with the properties of self-stimulation.

Ward (1959) did a study to determine which properties of the electrical stimulus were crucial in determining the threshold for self-stimulation and the rate of responding from septal placements. Using a sine wave stimulus of constant current (.09 milliamperes) and duration (.9 second per press), he found

that the alternation frequency of the current had little effect on rate of pressing in the range from 10 to 1,000 cycles per second; above and below these frequencies the rate fell off sharply. In the threshold experiments he used a stimulus of .2 second duration in the form of biphasic rectangular pulses; the length, frequency, and intensity of which he varied systematically. If we regard the threshold as a constant  $C$ , his results can be summarized by the equation  $I^2t = C$ , in which  $I$  equals the intensity of the current and  $t$  equals the length of the pulse.<sup>3</sup> The number of pulses did not materially affect the threshold in the range from 20 to 100 pulses per second. At 10 pulses per second he could not get his rats to maintain responding.

Albino and Lucas (1962) claimed to have demonstrated that stimulation at one positive electrode outside the septum immediately prior to septal stimulation facilitates responding for the septal stimulation. The presentation of their data, however, makes it impossible to decide whether the data justify such a conclusion or not.

#### OTHER THEORIES

There is a paucity of rigorous theories attempting to deal with the complete range and peculiar characteristics of the results from self-stimulation experiments.

Several theorists, especially those who incline toward drive reduction or homeostatic theories of learning, have suggested ways in which the rewarding effect of the stimulation could be explained. *Prima facie*, this rewarding effect is difficult to account for in terms of drive reduction. Brady (1958); Miller (1961a); and Nielson, Doty, and Rutledge (1958) have all suggested that the electrical stimulus may reduce drive by temporarily obliterating the drive

stimuli in the central nervous system—or, in Miller's words:

[by] temporarily blocking out motivational sources of stimulation. Such a functional elimination of motivational stimuli might occur during unconsciousness produced by a petit mal seizure [p. 576].

Nielson et al. feel strengthened in this view by the fact that the current intensities required to produce self-stimulation are considerably above those required for perception as conditioned stimuli.

This suggestion does not explain why many placements which produce evident convulsions do not produce self-stimulation, nor does it explain why a rat that a few minutes ago was sleeping in its cage will evince such a determination to return to a self-stimulation lever which it has been pulled away from after being shaped to press it. In other words, this does not explain the drive arousing effects of the electrical stimulus. No one has attempted to account for the other peculiarities of self-stimulation behavior reviewed above on the basis of this suggestion.

Miller (1961a) has also suggested that:

In some cases, the reward phenomenon (produced by subcortical stimulation) may be analogous to addiction, with the effects of one stimulation tending to remove the unpleasant aftereffects of the preceding one [p. 577].

As Miller points out, this explanation would only apply in some cases, since it has been demonstrated that stimulation from some placements apparently has no aversive component (Brodie et al., 1960). Nor, on the face of it, can this suggestion account for the interaction between self-stimulation and the various natural drives.

Mowrer (1960) suggested that self-stimulation placements are in areas wherein:

the cessation of drive (which homeostatic theory identifies with reinforcement) becomes translated from mere stimulation termination over into a more positive, more "active" state

<sup>3</sup> I am indebted to C. I. Howarth for having pointed out this simple mathematical summary of the findings.

of some kind, and that it is the latter which, in the final analysis, is the agent that fixates response tendencies [p. 207].

This is not implausible as far as it goes, but it does not seem to account for much of the data beyond the bare fact that stimulation from such placements is rewarding.

In an apparent extension of his earlier theory (Olds, 1955), Olds (1958a) has proposed that neural traces of a behavior sequence are laid down in the nervous system by their mere occurrence. Then, if certain reward areas are electrically stimulated at the end of the sequence, they will give the trace in question "afferent control of the neural elements which provide for the release of organized traces into overt behavior." The behavior will reappear when "some adequate representation of the environmental stimuli which preceded the behavior . . . [reorganizes] the whole trace [pp. 259-260]."

In its present form the theory does not explain the many motivational properties of electrical stimulation in rewarding areas. For example, it does not seem able to explain why animals so frequently require "priming" at the start of a day's trials, nor does it explain the rapid "extinction" of the responses learned for electrical stimulation.

Of the theories so far proposed, the structural theory appears most adequate to account for the complete range of results from studies on the phenomenon of electrical self-stimulation. The success of the theory at this more physiological level is especially noteworthy, since it was originally based chiefly on behavioral data from maze and Skinner box studies.

#### REFERENCES

- ALBINO, A. C., & LUCAS, J. W. Mutual facilitation of self-rewarding regions within the limbic system. *J. comp. physiol. Psychol.*, 1962, 55, 182-185.
- BEER, B., & VALENSTEIN, E. S. Discrimination of tones during reinforcing brain stimulation. *Science*, 1960, 132, 297-298.
- BOWER, G. H., & MILLER, N. E. Rewarding and punishing effects from stimulating the same place in the rat's brain. *J. comp. physiol. Psychol.*, 1958, 51, 669-678.
- BRADY, J. V. The paleocortex and behavioral motivation. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958.
- BRADY, J. V. Temporal and emotional effects related to intracranial electrical self-stimulation. In E. R. Ramey & D. S. O'Doherty (Eds.), *Electrical studies on the unanesthetized brain*. New York: Hoeber, 1960.
- BRADY, J. V., BOREN, J. J., CONRAD, D. G., & SIDMAN, M. The effect of food and water deprivation upon intracranial self-stimulation. *J. comp. physiol. Psychol.*, 1957, 50, 134-137.
- BRADY, J. V., & CONRAD, D. G. Some effects of limbic system self-stimulation upon conditioned emotional behavior. *J. comp. physiol. Psychol.*, 1960, 53, 128-137.
- BRODIE, D., MORENO, O. M., MALIS, J. L., & BOREN, J. J. Nonreversibility of the appetitive characteristics of intracranial stimulation. *Amer. J. Physiol.*, 1960, 199, 707-709.
- DELGADO, J. M. R., & ANAND, B. K. Increase in food intake induced by electrical stimulation of the lateral hypothalamus. *Amer. J. Physiol.*, 1953, 172, 162-168.
- DELGADO, J. M. R., ROBERTS, W. W., & MILLER, N. E. Learning motivated by electrical stimulation of the brain. *Amer. J. Physiol.*, 1954, 179, 587-593.
- DEUTSCH, J. A. *The structural basis of behavior*. Chicago: Univer. Chicago Press, 1960.
- DEUTSCH, J. A., & HOWARTH, C. I. Evocation by fear of a habit previously learned for intracranial electrical stimulation. *Science*, 1962, 136, 1057-1058.
- DEUTSCH, J. A., HOWARTH, C. I., BALL, G. G., & DEUTSCH, D. Threshold differentiation of drive and reward in the Olds effect. *Nature*, 1962, 196, 699-700.
- HODOS, W., & VALENSTEIN, E. S. Motivational variables affecting the rate of behavior maintained by intracranial stimulation. *J. comp. physiol. Psychol.*, 1960, 53, 502-508.
- HOEBEL, B. G., & TEITELBAUM, P. Hypothalamic control of feeding and self-stimulation. *Science*, 1962, 135, 375-377.
- HOWARTH, C. I., & DEUTSCH, J. A. Dissipation of a drive process as the cause of apparently fast extinction in ESB habit. *Science*, 1962, 137, 35-36.
- LARSSON, S. On the hypothalamic organization of the nervous mechanisms regulating

- food intake. *Acta physiol. Scand.*, Stockholm, 1954, 32(Suppl. No. 115).
- MARGULES, D. L., & OLDS, J. Identical "feeding" and "rewarding" systems in the lateral hypothalamus of rats. *Science*, 1962, 135, 374-375.
- MILLER, N. E. Experiments on motivation. *Science*, 1957, 126, 1271-1278. (a)
- MILLER, N. E. Objective techniques for studying motivational effects of drugs on animals. In S. S. Garattini & V. Ghatti (Eds.), *Psychotropic drugs*. New York: Elsevier, 1957. (b)
- MILLER, N. E. Implications for theories of reinforcement. In D. E. Sheer (Ed.), *Electrical stimulation of the brain*. Austin: Texas Univer. Press, 1961. (a)
- MILLER, N. E. Learning and performance motivated by direct stimulation of the brain. In D. E. Sheer (Ed.), *Electrical stimulation of the brain*. Austin: Texas Univer. Press, 1961. (b)
- MOWRER, O. H. *Learning theory and behavior*. New York: Wiley, 1960.
- NIELSON, H. C., DOTY, R. W., & RUTLEDGE, L. T. Motivational and perceptual aspects of subcortical stimulation in cats. *Amer. J. Physiol.*, 1958, 194, 427-432.
- OLDS, J. Physiological mechanisms of reward. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1955*. Lincoln: Univer. Nebraska Press, 1955. Pp. 73-139.
- OLDS, J. Runway and maze behavior controlled by basomedial forebrain stimulation in the rat. *J. comp. physiol. Psychol.*, 1956, 49, 507-512.
- OLDS, J. Adaptive functions of the paleocortex. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958. (a)
- OLDS, J. Effects of hunger and male sex hormone on self-stimulation of the brain. *J. comp. physiol. Psychol.*, 1958, 51, 320-324. (b)
- OLDS, J. Satiation effects in self-stimulation of the brain. *J. comp. physiol. Psychol.*, 1958, 51, 675-678. (c)
- OLDS, J. Self-stimulation of the brain. *Science*, 1958, 127, 315-324. (d)
- OLDS, J. Approach-avoidance dissociations in the rat brain. *Amer. J. Physiol.*, 1960, 199, 965-968. (a)
- OLDS, J. Differentiation of reward systems in the brain by self-stimulation techniques. In E. R. Ramsey & D. S. O'Doherty (Eds.), *Electrical studies on the unanesthetized brain*. New York: Hoeber, 1960. (b)
- OLDS, J. Differential effects of drives and drugs on self-stimulation at different brain sites. In D. E. Sheer (Ed.), *Electrical stimulation of the brain*. Austin: Texas Univer. Press, 1961.
- OLDS, J., KILLAM, K. F., & EIDUSON, S. Effects of tranquilizers on self-stimulation of the brain. In S. Garattini & V. Ghatti (Eds.), *Psychotropic drugs*. New York: Elsevier, 1957.
- OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain. *J. comp. physiol. Psychol.*, 1954, 47, 419-427.
- OLDS, J., & PERETZ, B. A motivational analysis of the reticular activating system. *EEG clin. Neurophysiol.*, 1960, 12, 445-454.
- OLDS, J., TRAVIS, R. P., & SCHWING, R. C. Topographic organization of hypothalamic self-stimulation functions. *J. comp. physiol. Psychol.*, 1960, 53, 23-32.
- REYNOLDS, R. W. The relationship between stimulation voltage and hypothalamic self-stimulation in the rat. *J. comp. physiol. Psychol.*, 1958, 51, 193-198.
- ROBERTS, W. W. Both rewarding and punishing effects from stimulation of posterior hypothalamus of cat with same electrode at same intensity. *J. comp. physiol. Psychol.*, 1958, 51, 400-407.
- SEWARD, J. P., UYEDA, A. A., & OLDS, J. Resistance to extinction following cranial self-stimulation. *J. comp. physiol. Psychol.*, 1959, 52, 294-299.
- SEWARD, J. P., UYEDA, A. A., & OLDS, J. Reinforcing effect of brain stimulation on runway performance as a function of interval between trials. *J. comp. physiol. Psychol.*, 1960, 53, 224-227.
- SIDMAN, M., BRADY, J. V., CONRAD, D. G., & SCHULMAN, A. Reward schedules and behavior maintained by intracranial self-stimulation. *Science*, 1955, 122, 830-831.
- SMITH, O. A. Stimulation of lateral and medial hypothalamus and food intake in the rat. *Anat. Rec.*, 1956, 124, 263-264.
- STEIN, L. Secondary reinforcement established with subcortical stimulation. *Science*, 1958, 127, 466-467.
- WARD, H. A. Stimulus factors in septal self-stimulation. *Amer. J. Physiol.*, 1959, 19, 774-782.
- ZEIGLER, H. P. Electrical stimulation of the brain and the psychophysiology of learning and motivation. *Psychol. Bull.*, 1957, 54, 363-382.

(Received June 4, 1962)

## ADAPTATION

IRENE MARTIN

*Institute of Psychiatry, London*

The various meanings of the term adaptation are examined, from the evolutionary context to the physiological. Of the 2 main uses in behavioral studies the 1st refers to the learned adjustments made in the organism's lifetime, and the 2nd to processes of response decrement, as in sensory adaptation. The former are judged adaptive by certain evaluative criteria, whereas the latter can be experimentally analyzed. These uses are considered broadly in relation to the categories of ontogenetic and phylogenetic adaptations, and specifically in relation to current views on learning.

The evolution of life has not been random, but oriented; and the major nonrandom orienting factor in this process is reasonably indentified as adaptation. Simpson (1958) therefore asserts that accounting for adaptation is the one central, inescapable problem of biology. But within the different biological sciences the evolutionary use of adaptation is modified and diluted: the term takes on various meanings and may cover many diverse kinds of processes.

Used loosely, adaptation refers to an enormous array of phenomena, ranging across a time span of millions of evolutionary years to a few minutes in the case of "adaptation" of sensory processes. In psychology alone, many references are made to adaptive behavior, processes of adaptation, habituation, and accommodation. Is there anything in common between these terms, or between these and the evolutionary study of adaptation?

It is time to inquire into this concept from a behavioral orientation. The purpose of this paper is, therefore, to survey the biological meaning of adaptation, first from an evolutionary and then from a behavioral point of view; next, to consider some experimental studies of so-called adaptation processes in psychology and physiology. Finally, the use of terms will be examined, and some

implications for psychological thought discussed.

### *Adaptation and Evolution*

It is an essential feature of evolutionary theory that changes occur only when and only as long as they are adaptive to the environment, and the unraveling of this idea represents one of the most fascinating and baffling problems of modern analytical biology. In this analysis, four major factors have been recognized: the nature of the external environment; development of morphological characteristics and physiological organization; behavior of the organism in such activities as food seeking, defense, and reproduction (especially in relation to natural selection); and genetics—the vast pool of genetic alternatives available in a population. These factors do not act singly, but interact in many complex ways.

The development of genetics over the past decades has made a significant contribution to current thought. We are now sufficiently free from Aristotelian teleology and well versed in genetics not to think of biological end states in terms of intelligent designers, but rather to acknowledge the *opportunistic* nature of adaptation; to see it as a historical process contingent upon the interplay of physical, ecological, and constitutional factors. In this process, dif-

ferent mechanisms such as the gorgeous plumage of birds, crests, colorings, and odors can all come to serve the same biological end of sexual attraction. Pittendrigh (1958) expressed this particularly lucidly, describing adaptation as a

patchwork of makeshifts pieced together, as it were, from what was available when opportunity knocked, and accepted in the hindsight, not the foresight of natural selection [p. 400].

Modern theories have shifted the emphasis away from Darwin's "struggle for survival." Although there is more or less unanimous agreement nowadays that his principle of natural selection is fundamental in evolutionary change, the emphasis is not so much on *individual* survival as on the successful reproduction and adaptation of *populations* (Simpson, 1949, 1958).

Increasing reproductive efficiency is one of the ways in which evolutionary progress, or improved *quality* of adaptations, can be demonstrated. The whole notion of evolutionary progress is, however, highly speculative, and it is uncertain what criteria should be applied. The various current ideas on specialization of adaptations, progressive independence from the environment, the ability to cope with a greater variety of environmental factors, etc., have been treated in many excellent texts, most notably by Herrick (1956) and Simpson (1953).

### *Adaptation and Behavior*

Evolutionary theories are largely based upon historical reconstructions of earlier forms of life. In the ontogenetic context, the term adaptation refers to current adjustments. Nevertheless, certain principles common to both phylogeny and ontogeny can still be discerned. Both refer to end directed systems (however ill defined this phrase might be),

both accept that there are alternative means of realizing or attaining their goals, and both set up certain criteria of adaptation such as successful adjustment to changing environmental conditions. A few thinkers, like Sommerhoff (1950), take the extreme view that *all* adaptations incorporate the same kinds of principles, and therefore can be treated within the same general formula.

There seem, however, to be certain clear-cut points of difference. The environmental background to which the individual adjusts in ontogenetic adaptations may be physical or social. In any event the adaptation is specific to the individual and is not genetically transmitted. Its medium is the central nervous system, whereas in phylogeny it is genetically determined.

The truth of the matter would seem to be more complicated than either extreme. On the one hand it is difficult to see how Sommerhoff's formula can encompass all the enormous variety of adaptation processes and ever be quantitative. On the other hand, some continuity, some interaction, clearly occurs between phylogenetic and ontogenetic adaptations.

The term "adaptive behavior" in the behavioral context is far less frequent and far less meaningful than it is in evolutionary thought. One reason is that the preoccupation with functional relations and general laws of learning necessarily demands a strictly controlled environment which allows the animal little opportunity for demonstrating or developing new molar patterns of behavior.

Most learning theories admit (albeit cursorily) that behavior is based on inherited adaptation patterns which have developed on the pragmatic basis of survival, and that it is their association with other stimuli and their satisfaction which are primary factors in shaping behavior. It is further recognized that

many of the drives which influence behavior are "acquired" as a result of individual experience; and even that increased *competence* in dealing with the environment can be rewarding (White, 1959).

While the capacity for learning is inherited, its actual exercise is entirely dependent upon the organism's interaction with its environment and the opportunities provided by it. Lack of opportunity can result in behavior which is undeveloped or "inappropriate"; conversely, a rich and intensely varying environment can tax the "adaptive" qualities of the organism and render its behavior "maladaptive" or "maladjusted." In this way, evaluative statements are built up such as "adjustment is an efficient relationship between an individual and his environment which enables him to live comfortably and without strain and conflict."

These views both involve the notion of criteria: (a) that biological adaptations such as eating and mating behavior can become more efficient or appropriate as a result of experience, and (b) that some forms of learned behavior (e.g., social) can be better adapted to the existing conditions than others.

*Evaluative Criteria of Adaptive Behavior.* These, like the ones mentioned earlier in connection with the quality of phylogenetic adaptations, are also highly speculative. To a certain extent, "efficiency" of eating and mating can be described in reasonably objective terms such as speed, directness of response, etc., where an individual's performance can be judged in relation to a behavioral norm or mean. This norm, combined with the experimenter's knowledge of the goal of the animal—escape from shock or procurement of food—leads somewhat inevitably towards a concept of behavior which is more or less adaptive. But evaluative criteria which refer

to the competence or efficacy of, for example, human social behavior, are extremely difficult to specify.

In one widely used sense, maladaptive behavior is that which appears as an extreme variant of a population norm. But the term has other connotations. One of the most persistent involves the notion of energy balance and expenditure.

*Concept of Energy in Adaptive Behavior.* It has often been specifically stated that an essential feature of adaptation is the delimitation of energy output to the minimum capable of meeting the requirements of a situation; and Thorpe (1956), for example, expresses the view that habituation is "economical" in that it saves the animal's energies. This links up with the view that maladaptive behavior results in a depletion of the organism's energy resources, and more generally with the common belief that normal behavior represents a kind of "biological equilibrium" or "energy balance" (Freeman, 1948).

Such concepts must be used with caution, for it is tempting to ascribe to behavioral energy the properties of physical energy; in the resulting confusion, more analytical studies of behavior may be hindered (Hinde, 1960a).

Energy models play a less obvious part in physiological theories, but there are many implicit references to energy processes. Cannon's emergency theory, for example, was fundamentally concerned with the redistribution of bodily resources to make energy rapidly available for muscular effort, and Selye's (1956) theory of the stress response and "adaptation energy" involves the belief that no living organism can be maintained continuously in a state of alarm because the drain upon internal energy resources is too great for it to cope with.

It seems hardly necessary to labor the

point that psychological factors can alter the pattern of behavior without any known change in hypothetical stores of energy (Kubie, 1947). We are too far from understanding general or adaptive behavior in terms of molecular patterns of "energy" changes in the organism to engage in sharp controversy over this point. But Hinde's warning must stand: there is no easy way out of analyzing behavior's intricacies by adopting facile models of any kind.

### *Adaptation as Response Decrement*

The next few sections will deal with some of the processes, labeled adaptation processes, which have been studied in detail; it will be noted that some of these processes are little, if at all, modified by learning—others considerably so.

*Reflex Adaptation.* Many kinds of responses have been studied over a wide range of animals (see reviews by Harris, 1943; Thorpe, 1956) from the now classical studies of Humphrey (1933) on the land snail and of Dodge (1931) on the knee jerk, eyeblink, and nystagmus responses, to more recent work on abdominal, biceps, and patellar reflexes (Hagbarth & Kugelberg, 1958; Lehner, 1941) and on the pupillary reflex to light (Cuppers, 1954; Lowenstein & Loewenfeld, 1951).

Many workers have studied the habituation of the widespread behavioral autonomic, muscular, and electroencephalogram components of orienting, alarm, or startle responses (Davis, Buchwald, & Frankmann, 1955; Dykman, Reese, Galbrecht, & Thomasson, 1959; Hinde, 1960b; Landis & Hunt, 1939; Scholander, 1961). The immediate autonomic and skeletal muscle reaction to a stimulus is diffuse, but after only a few repetitions of the stimulus this activity becomes less widespread, lower in amplitude, and shorter in dura-

tion until eventually there may be a complete failure to respond. In such cases, it seems as if the animal has learned not to respond.

There are also many recent electrophysiological studies on adaptation processes (Hernandez-Peon, Scherrer, & Velasco, 1956; Sharpless & Jasper, 1956). All agree on the kinds of factors which affect habituation, namely: (a) on the stimulus side—stimulus modality, intensity, duration, and inter-stimulus interval; and (b) on the response side—within-organism factors such as level of wakefulness or fatigue, intake of drugs, and previous adaptation level. Recovery of the habituated response can occur under certain conditions which include the lapse of time (spontaneous recovery), and the introduction of an extraneous stimulus or any sudden change in the total experimental situation (disinhibition).

*Sensory Adaptation.* All receptors manifest properties of adaptation, which means that a constant repetitive stimulus set off a series of receptor responses which gradually slows in time and eventually ceases. But the term adaptation in sensory functions has many meanings (Jenkins, 1951). For example, it may refer to the "physiological zero" of the subjective temperature scale, a kind of frame-of-reference concept which has been developed by Helson (1948) in his "adaptation-level" theory of perception. Or, as in the case of dark adaptation, it may refer to a process of recovery, i.e., increasing sensitivity.

The results of sensory research show that with careful control both of stimulus characteristics (e.g., intensity, duration, and wave length of light) and of receptor conditions (pupil size, retinal localization, etc.) rather regular and reproducible adaptation curves can be obtained. A number of restricted laws and theories have been developed to ac-

count for specific forms of sensory adaptation, such as the well-known photochemical theories of visual adaptation (Hecht, 1937).

There are no quantitative laws which can be generally applied to all forms of sensory adaptation—understandably so in view of the highly specific nature of receptors. From time to time, however, generalizations about sensory adaptation appear in the literature. One is that the temporal curve of adaptation is negatively accelerated, proceeding rapidly at first and gradually slowing; another is that the time required for adaptation increases with the intensity of the adapting stimulus; a third that rate of adaptation decreases as the number of receptors stimulated increases (Osgood, 1953).

It should be emphasized, however, that there exist not one but a great many adaptation curves of different shape and level, depending upon the conditions under which they were obtained. This comment should not be taken as implying that generalizations of this kind are unattainable, rather that they must be accompanied by some statement as to the range of conditions under which they hold.

#### *Differentiation of Adaptation Processes*

It seems to be commonly accepted that the categories of adaptation and habituation are different, although there have been isolated suggestions that some neural factor, such as increased threshold of synaptic transmission, is basically responsible for all types of ontogenetic adaptation processes. However, Humphrey (1933) and Harris (1943) both concluded after reviewing response decrement from *Protista* to man that habituation is found in organisms with presynaptic and synaptic nervous systems, and in those with no nervous systems at all!

Attempts to discriminate adaptation from habituation have been based on several factors: first, the *location* of events within the central nervous system. It is sometimes suggested that after excluding peripheral processes such as receptor and effector adaptation a "true" kind of habituation is reached which is distinctively central (Harris, 1943; Jenkins; 1951, Thorpe, 1956). Such views are ineluctably related to a somewhat outmoded reflex-circuit view of central nervous system organization, whereas current neurophysiological research points consistently towards extensive modification of incoming signals at every level from the receptor upwards (Hernandez-Peon et al., 1956; Livingston, 1959; Magoun, 1958).

A second attempt at differentiation has been based on time; for example, Harris (1943) argued that

certain stimulus-response integrations, activated at intervals long enough to preclude loss of ability of either receptor, conductor, or effector to respond maximally, will still decrement to zero level. To such decrement, uncomplicated by interaction with other integrations, we apply the term habituation [p. 387].

Such arguments have little force today. Many recent behavioral studies which deal with generalized responsiveness to stimuli, i.e., spatiotemporal patterns of electrical changes throughout the brain and widespread autonomic and muscular effects, show how the different elements of the total response are differentially habituated in time, some adapting rapidly and others more slowly. The reasons for this are not yet clear, but there are probably extensive and complex interactions between many of the response elements.

These arguments for discriminating adaptation and habituation processes are basically concerned with learned versus nonlearned responses. Thorpe (1956), for example, refers to habituation as a simple learning not to respond to stimuli

which tend to be without significance in the life of the animal. Others, e.g., Humphrey (1933), Bradley and Key (1958), also regard habituation as negative learning. It will be argued in the next section that truly adaptive behavior is that acquired in the experience of the individual and thus reflects a kind of learning process, and that nonlearned patterns of adaptation (such as sensory functions) are inherited evolutionary adaptations. No useful purpose can be found for distinguishing habituation as an independent term from adaptation.

#### TERMINOLOGY

Such a widely used term as adaptation must not, to quote Quine (1936), be allowed to rest upon an uncritical assumption of mutual understanding. It seems important, therefore, to establish some kind of classification of adaptation processes if only to aid communication and avoid the present confusion which attaches to the term.

Speaking from an evolutionary viewpoint, Simpson (1958) and Pittendrigh (1958) are both of the opinion that adaptation can refer to two broad classes, either to the *processes* or *end results* of processes which bring about and maintain an organism-environment relationship useful to individual organisms and populations.

The essential difference between these two groups is that the first are ontogenetic processes which occur within the organism and are not transmitted genetically to offspring. Pittendrigh includes in this group such processes as general homeostasis, habituation, antibody formation, and learning, i.e., practically all forms of what he calls "somatic adaptability."

He believes that the second group—end results of processes—comprises well-defined anatomical structures of claw, bill, tooth, or limb; behavioral features of the organism which serve proximate

ends such as food seeking, mating, and attack; and even some general features such as learning capacity. These phylogenetic adaptations evolve through very long periods of time until they reach a point of stability and little change, which also tends to last for appreciable periods of time. This way they may outdate their original purpose, as we can observe in many of man's adaptations which carry over from a more primitive and pungent existence.

There is of course a transitional group between these two large groups to include adaptation processes of Group 1 which are in course of becoming Group 2 adaptations: differential reproduction may favor a particular characteristic so that it eventually results in a change in the genotype. This is a historical process which would span many generations.

This simple classification into ontogenetic and phylogenetic adaptations is inadequate from a behavioral orientation. It is pointless to place antibody formation in the same class as learning, which to the psychologist introduces new and plastic properties of behavior such as long-term modification of responsiveness, opportunities for making varied and alternative responses to reach goals and objects, and the acquisitions of new goals not obviously related to biological needs.

Those processes such as antibody formation, receptor functions, etc., which are not subject to much modification as a result of experience, or, especially, of learning, should not be placed in the first group; they belong instead to the second group of phylogenetic adaptations. These are usually studied in historical perspective—inferred from fossils of bygone ages which provide a succession of "stills" of life. But some of them are available for study in living organisms.

For example, comparative studies

point to the incredible variety of photo-receptors to be found among animals, ranging from diffuse photosensitivity of the whole body through to the fully developed image forming eye with lens, iris, and other complexities (Simpson, 1953). Each is characteristically useful in a particular kind of environment and way of life, and together they represent multiple solutions of adaptational problems. In the experimental context they can be studied as phylogenetic, not ontogenetic, adaptations. To ask "are these visual processes adaptive?" is to ask a question of evolutionary biology, not of psychology or physiology.

Some of the multidimensional factors which are involved in the analysis of adaptation can be summarized as follows:

1. Ontogenetic adaptation processes: the new adjustments and responses the organism makes as a result of its own learning and experience. These changes are (a) not necessarily common to all members of the population; they may remain restricted to individuals. (b) They are not genetically transmitted. The medium of their occurrence is the central nervous system; they may be conveyed to other individuals, however, through social-cultural media. (c) They are of a short-term stability compared with evolutionary adaptations; they can result in relatively stable and persistent habits of the organism, and thus refer to a kind of behavioral steady state.

2. The end results of adaptation processes—phylogenetic adaptations: examples are limb structure, physiological systems, and patterns of behavior such as escape, defense, etc. These (a) mediate biological ends—survival and reproduction of the population; (b) are common to the population; (c) are genetically transmitted; and (d) are stabilized patterns of organization which often persist indefinitely.

#### IMPLICATIONS FOR PSYCHOLOGY

The term adaptation has two main uses in psychology. The more common refers to response decrement through repeated elicitation. The second use is more evaluative (and applied to behavior based on a combination of response increment and decrement). For example, when a behavior pattern develops to some specific environmental demand it is judged to be adaptive or maladaptive on the basis of a number of nonquantifiable concepts.

From this complex we derive certain impressions as to what constitutes well-adapted behavior which as yet can hardly be formulated into any precise experimental program. For in the first place they deal with complex and molar *patterns* of behavior resulting from "real-life" organism-environment interaction, and in the second place they frequently involve the content rather than the process of learned behavior.

The notion of adaptation as response decrement is quite different. It is a more molecular, more easily isolated behavioral process which can be experimentally investigated. There is clearly a very strong case for avoiding confusion by insisting that the neutral term "response decrement" replace the more common adaptation and habituation in the psychological literature. Moreover, it would help establish continuity between different experimental approaches to response decrement and increment, and highlights a point recently made by Harlow (1959) that there exists no fundamental difference other than complexity, between sensory adaptation, habituation learning, conditioned response learning, and the kinds of learning described as reasoning and thinking. Looked at in this way, the study of sensory response decrement, which in some areas is so advanced and quantitative,

could represent a valuable methodological paradigm.

For all the lip service paid to biological adaptations—drives, needs, innate mechanisms, etc.—psychologists in recent years have directed little attention to instinctive behavior as such. It is the ethological group which has become increasingly concerned with the analysis of instincts, and thus with more highly integrated sequences of behavior. Ethology has far more immediate affinities with the biological significance of behavior than has psychology; but there are signs of change. Breland and Breland (1961) after running into difficulties in conditioning animals comment:

After 14 years of continuous conditioning and observation of thousands of animals, it is our reluctant conclusion that the behavior of any species cannot be adequately understood, predicted, or controlled without knowledge of its instinctive patterns, evolutionary history, and ecological niche [p. 684].

Perhaps their difficulties center on the question of what is an adequate explanation or understanding of behavior. They were struck by the sort of behavior their animals demonstrated; most psychological experimenters ignore the content of learning and concentrate on the "how," and in this respect they have been hopeful of establishing general laws of wide applicability. These are two complementary, equally valid approaches; the Brelands' findings are not incompatible with either, but have led them to the limits of certain general laws. We are reminded once more not only that drive and learning theories will one day be trimmed down to a sensible size, but that they will almost certainly be better shaped to take account of innate phylogenetic adaptation patterns.

Learning theories will undoubtedly also benefit from a closer look at the nature of the response decrement to the unconditioned stimulus which occurs in

many classical conditioning situations. Although it is true that some reflexes modify little upon repeated stimulation, many alter considerably. Indeed, in the case of some responses—the galvanic skin response, electroencephalogram blocking, electromyograms, and the abdominal skin reflex, for example—response decrement is an obtrusive factor, and hand in hand with this goes difficulty in achieving or maintaining the conditioned response (if conditioned it is).

These factors have led to debates not only on what specific criteria should be applied to classical conditioning, but on the more general problem of the physiological measurement of drive. The galvanic skin response not only rapidly diminishes to successive stimuli but reappears if the stimulus situation is changed in any way, an erratic pattern which is reflected in the difficulty of establishing a conditioned response which excludes sensitization and pseudoconditioning (Stewart, Stern, & Winokur, 1961). This raises questions about traditional definitions of classical conditioning, and whether pairing of the conditioned stimulus and the unconditioned stimulus can or should be given a status different from other processes of response increment (see Hagbarth & Kugelberg, 1958).

Further, when both the specific response being conditioned and the general pattern of physiological and behavioral activity modify upon repeated application of the unconditioned stimulus, it seems reasonable to infer that the organism's level of drive is also changing throughout the experiment. These physiological data may be difficult to incorporate into unitary behavioral concepts of drive, but should provide more direct links with the neurophysiological approach to learning and conditioning (Malmo, 1958).

Thus there are some broad and some specific implications for psychology arising from a review of the multidimensional approach to biological adaptations. One is obvious: that a preoccupation with general behavioral laws and theory should not preclude a wider sympathy for other less systematic approaches in biology; that interdisciplinary research with psychology's near neighbors—ethology and genetics—must be fruitful; that the biological significance of behavior, albeit complicated and unwieldy from an experimental point of view, must sooner or later be tackled experimentally.

More specifically, studies of response decrement show it to be an obtrusive factor, especially in classical conditioning. In some response modalities, conditioning seems only to be one of several processes of response facilitation which can be offset by opposing processes of response decrement. These considerations should lead to a reanalysis of drive and arousal concepts in current learning theory.

To conclude, the pattern of alarm, alerting or startle reaction which occurs to unexpected and novel stimuli, can be used to illustrate in summary form the different aspects of adaptation.

The elements of this *phylogenetic* adaptation pattern can be studied comparatively so that its content and expression can be related to the type of life the animal leads (or has led) and the ecological niche which it occupies. Thus one animal will be alerted by movement, another by a particular shape, and so on. In addition, the modifications which occur in the overall pattern to repetitive stimuli as a result of learning and experience can be studied experimentally as processes of *response decrement*. Undoubtedly some elements (e.g., the blink of the startle response, the effects of certain "sign stimuli")

remain undiminished whereas other aspects of the response pattern are rapidly inhibited. These processes of response modification are apparently systematic and can be expressed lawfully. Lastly, the relatively stable state of *ontogenetic* adaptation reached by the organism after considerable experience, such that the alerting response may not occur at all to certain stimuli, can be systematically examined, and proposed *criteria* for assessing whether this behavior is adaptive or maladaptive made more explicit.

## REFERENCES

- BRADLEY, P. B., & KEY, B. J. The effect of drugs on arousal responses produced by electrical stimulation of the reticular formation of the brain. *EEG clin. Neurophysiol.*, 1958, 10, 97-110.
- BRELAND, K., & BRELAND, MARIAN. The misbehaviour of organisms. *Amer. Psychologist*, 1961, 16, 681-684.
- CUPPERS, C. Continuous recording of the direct and consensual pupillary reflex. *Graefes Arch. Ophthalmol.*, 1954, 155, 588-616.
- DAVIS, R. C., BUCHWALD, A. M., & FRANKMANN, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychol. Monogr.*, 1955, 69(20, Whole No. 405).
- DODGE, R. *Conditions and consequences of human variability*. New Haven: Yale Univer. Press, 1931.
- DYKMAN, R. A., REESE, W. G., GALBRECHT, C. R., & THOMASSON, PEGGY J. Psychophysiological reactions to novel stimuli: Measurement, adaptation, and relationship of psychological and physiological variables in the normal human. *Ann. N. Y. Acad. Sci.*, 1959, 79, 43-107.
- FREEMAN, G. L. *The energetics of human behavior*. Ithaca: Cornell Univer. Press, 1948.
- HAGBARTH, K. E., & KUGELBERG, E. Plasticity of the human abdominal skin reflex. *Brain*, 1958, 81, 305-318.
- HARLOW, H. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. *General systematic formulations, learning, and special processes*. New York: McGraw-Hill, 1959. Pp. 492-537.
- HARRIS, J. D. Habitatory response decre-

- ment in the intact organism. *Psychol. Bull.*, 1943, 40, 385-422.
- HECHT, S. The nature of the visual process. *Harvey Lect. Ser.* 33, 1937, 35-64.
- HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.
- HERNANDEZ-PEON, R., SCHERRER, H., & VELASCO, N. Central influences on afferent conduction in the somatic and visual pathway. *Acta neurol. Latin-Amer.*, 1956, 2, 8-22.
- HERRICK, C. J. *The evolution of human nature*. Austin: Univ. Texas Press, 1956.
- HINDE, R. A. Energy models of motivation. *Symp. Soc. Exp. Biol.*, 1960, 14, 199-213. (a)
- HINDE, R. A. Factors governing the changes in strength of a partially inborn response, as shown by the mobbing behaviour of the chaffinch (*Fringilla coelebs*): III. The interaction of short-term and long-term incremental and decremental effects. *Proc. Roy. Soc., Ser. B*, 1960, 153, 398-420. (b)
- HUMPHREY, G. *The nature of learning*. New York: Harcourt Brace, 1933.
- JENKINS, E. L. Somesthesia. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1172-1190.
- KUBIE, L. S. The fallacious use of quantitative concepts in dynamic psychology. *Psychoanal. Quart.*, 1947, 16, 507-518.
- LANDIS, C., & HUNT, W. A. *The startle pattern*. New York: Farrar & Rinehart, 1939.
- LEHNER, G. A study of the extinction of unconditioned reflexes. *J. exp. Psychol.*, 1941, 29, 435-456.
- LIVINGSTON, R. B. Central control of receptors and sensory transmission systems. In J. Field (Ed.), *Handbook of physiology*. Vol. 1. *Neurophysiology*. Washington, D. C.: American Physiological Society, 1959. Pp. 741-760.
- LOWENSTEIN, O., & LOEWENFELD, I. E. Types of central autonomic innervation and fatigue: Pupillographic studies. *Arch. Neurol. Psychiat.*, 1951, 56, 580-599.
- MAGOUN, H. W. Non-specific brain mechanisms. In H. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univ. Wisconsin Press, 1958. Pp. 25-36.
- MALMO, R. B. Measurement of drive: An unsolved problem in psychology. *Nebraska symposium on motivation: 1958*. Lincoln: Univ. Nebraska Press, 1958. Pp. 229-265.
- OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univ. Press, 1953.
- PITTENDRIGH, C. S. Adaptation, natural selection and behaviour. In A. Roe & G. G. Simpson (Eds.), *Behaviour and evolution*. Yale Univ. Press, 1958. Pp. 390-416.
- QUINE, W. Truth by convention. In *Philosophical essays for Alfred North Whitehead*. New York: Longmans Green, 1936.
- SCHOLANDER, T. Studies in the habituation of autonomic response elements. *Acta Soc. Med. Upsaliensis*, 1961, 66, 33-42.
- SELYE, H. *The stress of life*. New York: McGraw-Hill, 1956.
- SHARPLESS, S., & JASPER, H. Habituation of the arousal reaction. *Brain*, 1956, 79, 655-680.
- SIMPSON, G. G. *The meaning of evolution*. New Haven: Yale Univ. Press, 1949.
- SIMPSON, G. G. *The major features of evolution*. New York: Columbia Univ. Press, 1953.
- SIMPSON, G. G. Behaviour and evolution. In A. Roe & G. G. Simpson (Eds.), *Behaviour and evolution*. New Haven: Yale Univ. Press, 1958. Pp. 507-535.
- SOMMERHOFF, G. *Analytical biology*. London: Oxford Univ. Press, 1950.
- STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, 68, 60-67.
- THORPE, W. H. *Learning and instinct in animals*. London: Methuen, 1956.
- WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychol. Rev.*, 1959, 66, 297-333.

(Received June 25, 1962)

## NEUROCHEMISTRY AND LEARNING

JOHN GAITO AND ALBERT ZAVALA

*Kansas State University*

Hypotheses of molar and molecular neurochemical approaches to learning are discussed. Molar approach equates learning with chemical changes of the synapse whereas molecular approach refers learning to changes in molecular structure of nucleic acids. 2 approaches are evaluated relative to various learning phenomena.

The topic of learning is a central problem for psychologists. The major learning approaches which have dominated psychology have been the treatments within a psychological frame of reference exemplified by Thorndike, Hull, Guthrie, Skinner, the Gestaltists, Tolman, and others. However, in recent years a number of other approaches have appeared on the scene, viz., mathematical, analogical, neurological, and neurochemical.

The mathematical (or symbolic) approach to learning which has generated much research in recent years has been of a probability or stochastic nature (e.g., Bush & Mosteller, 1955; Estes, 1950; Overall, 1960; Restle, 1955). These mathematical models have served as convenient guides and have provided valuable insights concerning learning phenomena.

In recent years psychology has been permeated by a rash of models in which different types of analogies provide insight into the functioning of the organism. The main concern of some of these models has not been with learning; however, most of these analogies can include learning phenomena. These analogues include communication models (Miller, 1953), servomechanisms (Weiner, 1948), computer models (Resenblatt, 1958), mechanical models (Broadbent, 1957), and others. These models, like the mathematical approach, have provided interesting and ingenious ideas concerning behavior.

Both the mathematical and analogical approaches are of *extraorganismic* nature, i.e., they utilize external aspects to apply to the organism. On the other hand, the neurological and neurochemical approaches are *intraorganismic* in that they deal directly with the internal aspects of the organism.

The neurological approaches to learning have been numerous. The most influential ones have been those by Lashley (1929) and by Hebb (1949). The cell assemblies and phase sequences of the latter have permeated much of psychological thinking. Recently Pribram (1960) has presented an interesting theoretical approach in an excellent review. Other excellent reviews of recent research on learning phenomena include articles by Morrell (1961), John (1961), and Thomas (1962).

The neurochemical, or biochemical, approach to learning is a very recent one and appears to offer great promise for the future. In this area some very exciting research results have been obtained. We would like to review this approach.

The neurochemical approach to behavior may be said to have had its inception when Thudichum published a monumental treatise (Thudichum, 1884; cited by Page, 1955). Thudichum described the first systematic attempt to understand the chemical mechanisms of nervous tissue. After Thudichum's death in 1901 work on neurochemistry all but stopped. Ap-

proximately 35 years later, scientists again began to study the chemical mechanisms of nerve tissue.

In recent years this renewed interest in neurochemistry has been reflected by a number of results which have been pertinent to psychology. For example, Williams (1956) has investigated the biochemical make-up of humans and lower animals and has indicated that underlying the differences in behavior of organisms are differences in chemical activity. An exciting neurochemical approach is the attempt to identify specific chemicals such as serotonin, adrenolutin, acetylcholine (ACh), etc., as the causative agents for psychotic behavior (Rinkel & Denber, 1958; Rubin, 1959).

The neurochemical approach to learning is extremely new, having developed within the past decade. There have been two major treatments of learning within this approach, one remaining at a rather molar chemical level whereas the other is concerned with submolecular changes.

#### MOLAR CHEMICAL APPROACH

The molar chemical approach deals with the relationship of various chemicals such as ACh or calcium with the acquisition and retention of specific behavior indicative of learning. This neurochemical approach rests on the general hypothesis that variations in brain chemistry are major determinants of adaptive behavior in animals. The California group (Krech, Rosenzweig, and others) are the foremost proponents of this approach.

The main thesis of the California group is that changes in ACh are related to learning and adaptive behavior. They have conducted a series of experiments combining biochemical analytic techniques and psychological procedures. They have hypothesized that experience and training may significantly alter the

concentration of brain ACh and cholinesterase (ChE; Bennett, Rosenzweig, Krech, Karlsson, Dye, & Ohlander, 1958; Bennett, Krech, Rosenzweig, Karlsson, Dye, & Ohlander, 1958; Krech, Rosenzweig, Bennett, & Krueckel, 1954; Krech, Rosenzweig, & Bennett, 1956, 1959; Krech, Rosenzweig, Bennett, & Longueil, 1959; Rosenzweig, Krech, & Bennett, 1956, 1960). These individuals have maintained that transmission of nerve impulses is accomplished by the discharge of ACh from presynaptic neurons. Inactivation of ACh by ChE, as soon as the former stimulates the postsynaptic neuron, preserves discrete transmission of impulses.

The California group was interested in the relationship between ACh and behavior. However, no reliable techniques for the measurement of ACh were available. Therefore, they decided to use amounts of ChE as an index of ACh levels because the former was considered to be a stable component which is easily and reliably measured. This decision was a tentative one made with some misgivings.

The biochemical analytic procedures involved decapitating an animal, taking samples from desired portions of the brain (of rats), weighing the sample, and homogenizing the tissue in certain chemicals. After homogenizing the tissue, a number of complex chemical reactions were involved; the end result being a statement of the ChE activity per milligram of tissue.

The initial study by the group indicated that rats with strong spatial (right or left) preferences in maze running had greater ChE activity than rats with strong visual (light or dark) preferences (Krech et al., 1954). The greater ChE activity was present in somesthetic, visual, and motor areas of the cortex. These results were confirmed in a later experiment in which pentobarbital so-

dium was administered to depress ACh metabolism. Animals given the drug consistently showed visual preferences (Rosenzweig et al., 1956). Further results were in agreement with the first experiment (Krech et al., 1956).

A very interesting result was obtained when Bennett, Rosenzweig, Krech, Karlsson, Dye, and Ohlander (1958) studied the ChE activity of two strains of rats at various ages. It was found that descendants of the Tryon Maze-Bright strain showed greater average ChE activity than did descendants of the Tryon Maze-Dull strain for all ages investigated (29–527 days). The curves for both strains over days showed a rapid rise, reaching a peak at approximately 75–100 days, and then gradually decreasing. The two curves are remarkably similar to the curve of mental growth reported for humans (Wechsler, 1944). However, the similarity may be coincidental.

Another experiment indicated that subcortical ChE activity was negatively related to brain weight, but there was no relationship between cortical ChE and brain weight (Krech, Rosenzweig, & Bennett, 1959).

In a recent summary of their approach, Rosenzweig, Krech, and Bennett (1960) have rejected their hypothesis that ChE activity furnishes a good index of ACh activity. They maintained that each is under separate genetic control. However, their basic hypothesis that the ACh transmission system is intimately involved in learning was retained. This change was necessitated by results which showed that for some strains of rats high ChE levels were positively related to learning ability but in other strains there was a negative relationship. Further experiments suggested that the difference between ACh and ChE levels was of major importance, with greater learning ability being related to greater amount of ACh (within limits) at the synapse.

The experimental work conducted by Krech and associates lends credibility to their hypothesis. However, Tower (1958) has raised questions concerning the approach. He maintained that even though evidence favors ACh as the transmitter substance, the idea is controversial; ACh seems to be of less or no importance to certain systems, including much of the sensory functions upon which maze behavior is dependent; the approach does not differentiate between Type I and Type II ChE but only the former participates in ACh hydrolysis; ChE in the brain is far in excess of requirements, thus making difficult the uncovering of relationships with activity; the changes in ChE are so small as to be possible errors of sampling and analysis. Rosenzweig, Krech, and Bennett (1960) have countered this last criticism by indicating that even though the differences they have found are small, these have appeared consistently.<sup>1</sup>

Another approach related to the ACh-ChE hypotheses is that of Overton (1958, 1959a, 1959b). Overton's approach is very simply stated as a calcium displacement hypothesis: decrease in calcium concentrations at the synapse facilitates learning; increases in calcium retards learning. According to Overton, stimulation produces reverberatory circuits which are the basis for immediate memory. If the activity of these circuits continues as long as 30 seconds, permanent changes may occur at the synapse which will maintain permanent memory for the stimulating event. These changes at the synapse consist of the displacement of calcium by acetic acid,

<sup>1</sup> More recent research by the California group has handled more of these criticisms and they have reported further outstanding results: they found that environmental stimulation leads to an increase in cortical weight, in thickness of cortex, in both Types I and II ChE, and a tendency toward an increase in the ratio of glia to neurons in the cortex (Rosenzweig, 1963).

one of the by-products of the breakdown of ACh by ChE. Strong stimulation is supposed to produce greater concentrations of acetic acid which increases the probability of removing calcium deposits that occur naturally at the synapse.

Overton gives the results of experimentation which are consistent with his calcium displacement hypothesis. He fed two groups of rats diets of calcium, the diet of one group containing 35 times as much calcium as the other. The group on low calcium diet performed better in learning a maze than did the high calcium group. He also found that the high calcium diet group contained more calcium in brain neural tissue than did the low calcium group. Although these results tended to be consistent with his hypothesis, he suggested that the different diets may have had differential effects on sensory threshold and discriminative ability rather than on memory of the maze. He also suggested that altered metabolism of other electrolytes in the system brought about by abnormal calcium diets may have caused the performance differences. A third alternative he offered was that the calcium diets may have affected the rats' cage activity, producing differential rates of retroactive inhibition.

The approach of both Overton and the California group are remarkably similar. Overton merely considers synaptic events other than ACh-ChE interaction and attempts to develop a more involved theoretical schema. His overall approach reflects a merger of Hebbian and Hullian ideas.

Two major difficulties are posed for proponents of this approach. The central thesis of both Krech and his associates and of Overton is that the chemicals being measured operate at the synapse. Their techniques are not refined enough to conclusively indicate that this is so. Thus, the ACh-ChE and calcium activity could be more diffuse.

However, Tower (1958) maintains that ACh appears to be concentrated at synaptic endings and that the supply seems to vary with synaptic activity.

The second difficulty is one that confronted the Hebbian system (Hebb, 1949). Milner (1957) indicated that a deficiency of the Hebbian system was its emphasis on neural excitation to the exclusion of inhibitory phenomena and maintained that these latter events could not be explained by lack of facilitation or by invoking the principle of refractoriness. Hence, he introduced an active mechanism for inhibition involving an interaction between cells. The same criticism can be leveled at the molar chemical approaches. Both approaches appear to be handling inhibitory phenomena as a lack of facilitation.

#### MOLECULAR CHEMICAL APPROACH

The molecular or specificity neurochemical approach to the experiential code is not new. A number of people have expressed the idea that memory involves a molecular change in certain tissue, for example, Pauling and Weiss during the Hixon Symposium (Jeffress, 1951) and Gerard (1960). In a recent book on the nature of chemical bonding, Pauling (1960) concludes:

that thinking, both conscious and unconscious, and short-term memory involve electromagnetic phenomena in the brain, interacting with the molecular (material) patterns of long-term memory, obtained from inheritance or experience [p. 570].

A few years ago protein molecules were considered the most likely candidates to contain the memory trace. Katz and Halstead (1950) and Halstead (1951) favored the hypothesis that nucleoproteins were the substances which had the ability to act as templates on which replica molecules were formed. At first the neurons of the brain were supposed to contain random configurations of protein. Stimulation of neural

tissue by impulses caused the randomly oriented molecules to assume a specific configuration. Nucleoproteins were involved in these reorientations and became templates, thus providing a memory mechanism. They believed that these templates are like those of the germ cells in representing native endowment but differ from the latter in arising from external stimulation. They stated that the ordering of the protein templates could take place in various components of the cell and its processes, including the synapse. However, the reorganized protein replicas ultimately reside in the neural membrane where they participate as "traces."

Recently the protein molecule has been supplanted by the nucleic acids as the potential memory mechanism. In the last few years a number of individuals have independently suggested the hypothesis that deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) function as the molecular mechanism in which experiential events are symbolized (Dingman & Sporn, 1961; Gaito, 1961; Hyden, 1959).

Before launching into a discussion of these hypotheses, some background information concerning nucleic acids is in order.

It has been conclusively established by virologists and other microbiologists that DNA is the genic material (Butler, 1959; Crick, 1954, 1957; Lederberg, 1960; Sager & Ryan, 1961; Stanley & Valens, 1961; Strauss, 1960). DNA is a large double strand molecule wound in a helix, with each strand being the complement of the other. Each strand has a recurring sequence of a phosphate followed by deoxyribose sugar. The strands are attached at the sugars by means of bases called purines and pyrimidines. The former are the larger and consist of adenine and guanine; the latter, of thymine and cytosine. Thus, the only apparent difference in DNA

molecules is in the sequence of these bases. If we consider the permutations of the four bases, with the restriction that adenine must be linked to thymine and guanine to cytosine (a restriction which occurs in nature), there are only four possibilities. Even though there are only four possible bases, the strands of the molecule are extremely long and allow for many possibilities. DNA molecules consist of thousands of units (Kornberg, 1959). In the human it has been suggested that there are approximately 800,000 DNA molecules, each with 40,000 of these units (Stanley & Valens, 1961). But let us assume that only 1,000 units are present. In this case, the number of possibilities for a single DNA molecule is  $4^{1,000}$ , assuming independence of sites. Such figures are so enormous that they exceed the limits of comprehension.

If we assume that the direction of bonding (adenine-thymine is not different than thymine-adenine, for example) is not of importance, the number of possibilities would be  $2^n$ , where  $n$  refers to the number of independent units. Again the result is an astronomical figure.

However, we must allow for the possibility that the various sites or units are not independent of each other. This appears to be the case in that Chargaff (1958) has indicated that runs of three or more purines followed by three or more pyrimidines occur frequently. Such lack of independence would reduce the overall number of possibilities but the remaining number of possible events would still be tremendously large.

Thus it is obvious why microbiologists believe that these bases may furnish the "language" of the genes. Crick (1954) has likened the bases to the dots and dashes of the Morse code. He states that there is enough DNA in a single cell of the human body to encode about 1,000 large textbooks.

The other nucleic acid, RNA, is a single strand analogue of DNA. DNA has deoxyribose sugar and the bases, adenine, guanine, thymine, and cytosine; RNA has ribose sugar and three of the DNA bases but uracil instead of thymine.

DNA is found in the nucleus of the cell while RNA is located both in the nucleus (in the nucleolus) and in the cytoplasm of the cell. In the nerve cell, Hydén (1959) has indicated that 90% of the RNA is located in the soma of the cell; the remaining 10% is found in the dendrites. He was unable to find any measurable amounts in the axon. The RNA in the cytoplasm is mainly concentrated in the Nissl substance.

DNA and RNA are intimately involved in cellular functions which are of vital concern to the organism. For example, in protein synthesis it is accepted that DNA sets the process in motion by giving "instructions" to the RNA molecules which then interact with amino acid molecules to form proteins. Thus it is assumed that DNA passes the genetic code to the RNA molecule via the linear sequence of bases. Research results support this assumption.

A number of investigators have shown that the base ratio of DNA of a bacterial virus, T2, is analogous to that of its RNA. This suggests that the similarity may go further, even to a detailed correspondence of the bases in the two. That this correspondence is so was indicated by Hall and Spiegelman (1961). They used purified T2-RNA in order to provide opportunity for the RNA to combine with its DNA complement, unhindered by nonspecific interactions with irrelevant RNA. Their data demonstrated that molecules are formed consisting of T2-RNA and its homologous DNA. These results are strong evidence for the complementarity of the base sequences in RNA and DNA. They also suggested that the interaction between

DNA and RNA in which a strand of RNA is fashioned in the complementary image of DNA may be the means whereby the genetic code is transmitted to the RNA, which then directs the synthesis of proteins. The exact mechanisms involved in the synthesis and functioning of messenger RNA were suggested by Hurwitz and Furth (1962).

During the last 13 years, Holger Hydén (1959, 1961) of the University of Göteborg has been doing extensive research concerning RNA and proteins. Using very refined micro-molecular biochemical techniques, he stimulated animals and then analyzed the contents of nerve cells. His results demonstrated that RNA and proteins are produced in the nerve cells at a rate which follows neuronal activity. He maintained that the nerve cell fulfills its function under a steady and rapidly changing production of proteins, with the RNA as an activator and governing molecule. He hypothesized that memory involves a change in the sequence of bases in the RNA molecule through frequency modulation; one or more bases are exchanged with the surrounding cytoplasmic materials. The details of the formation of the memory trace are as follows:

1. Specification of RNA in the Nerve Cell: The first series of impulses generated in sensory cells or in motor neurons pre- or postnatally change the stability of one or more of the four bases of the RNA molecule at a certain site along the molecule. Precisely which changes in stability will occur depend on the modulated frequency of the first series of impulses generated in the nerve cell. This effects a change of one against another base from the surrounding pool. The new base at this space is now stable under the influence of the modulated frequency. The new pattern remains and constitutes the specification of the RNA in the nerve cell.

2. Formation of RNA Protein under

Neural Activity: Since the sequence of the bases in the template RNA is now changed, new protein formed through the mediation of the RNA will also be specified.

3. Dissociation of Protein: Stimulation causes a rapid dissociation of the specified protein and the combination of the dissociated products with a complementary molecule or an energy activation of the specified dissociated protein.

4. Activation of Transmitter Substance: Through a rapid combination of the dissociated protein with a complementary molecule in the third step, an activation occurs of the transmitter substance, and the postsynaptic structure is excited.

Hyden holds that the neurons are interconnected in complex tridimensional chains. Thus, when neurons in an already established memory trace are activated, regardless of how many there are, or where they are located, the hypothesis implies the reaction consists of the rapid third and fourth steps in the scheme. The impulses in the form of modulated frequencies activate the specified protein and release the transmitter substance causing the next neuron in the chain to be activated or excited. The transmitter substance may be ACh or any assumed inhibitory substance.

Hydèn apparently agrees with some type of Hebbian approach regarding the linking of cells to form chains or circuits in that he maintained that the reason that only neurons of a chain respond is because protein, once specified by a modulated frequency, responds to the same electrical pattern whenever repeated but will not react to other patterns.

In Hydèn's system strict localization of function in the brain is not required. The same neuron could participate as a link in many neuronal networks, depending on the response of the sequence of

the bases in the RNA to the modulated frequency. Association might grow out of successive specification of neurons.

We may summarize Hydèn's theory in less technical terms as follows. A nerve cell responds differentially depending on whether the pattern of impulses it receives is novel or familiar, as well as on the pattern itself. No protein will have the correct configuration if the incoming pattern of impulses is new; therefore no dissociation of the protein (followed by transmitter substance explosion) can occur. The electrical pattern must first shape a new RNA molecule, which in turn shapes a protein molecule that can dissociate. The molecule fragments and then reacts with a complementary molecule, causing the triggering of a substance (excitatory or inhibitory) across the synapse. If, on the other hand, the incoming impulse is familiar, protein molecules will already be present that are capable of dissociating rapidly. Each cell may perpetuate within itself a large number of unique patterns of RNA and protein. A giant RNA molecule could accommodate along its length many different sequences shaped by different impulse patterns that have coursed through the neuron.

Gaito (1961) suggested that changes might occur at the attachment of the two strands of DNA, with adenine at one locus changing to guanine and the associated pyrimidine changing from thymine to cytosine. Deletion, addition, and rearrangement of bases could be other means of changing the code. Thus, these changes would provide a basis for modification of the genetic potential in nerve cells by means of external stimulation, learning. This would be a somatic mutation which would not be transmitted to the offspring. However, other possible biochemical mechanisms were suggested which could be involved in memory functions, changes either in RNA or in the amino acids to subserve

memory functions rather than any changes in the DNA. The changes in DNA, RNA, or amino acid molecules could occur either in the nucleus or cytoplasm of the nerve cell. It is also possible that some of these modifications might occur in the cell processes, even at the synapse.

Dingman and Sporn (1961) hypothesized that RNA changes were the basis for memory. However, they suggested that the linear sequence of bases (primary structure) was only one possible means of codifying experiential events. They suggested that changes in the helical structure (secondary structure) and overall configuration (tertiary structure) could be the basis for memory.

Having looked at the hypotheses that some modification of DNA and/or RNA constitutes the basic change in learning, let us examine the empirical data relative to the nucleic acids. Evidence superficially seems to favor the RNA molecule rather than DNA, as a possible experiential code.

### DNA

Of the two nucleic acids, DNA is the most stable. The average DNA content appears to be relatively stable even during marked physiological alterations of cells whose nuclear size and protein content may vary widely (Alfert, 1957). There is general agreement concerning the absence of DNA turnover in rapidly growing cells and a limited rate of turnover in slowly growing cells or in resting bacteria (Sinsheimer, 1960). However, Sinsheimer reports a drop of 20% in DNA content in certain cells of the rat liver upon treatment with cortisone and an increase of 11% in the nuclei of the adrenal medulla of rats maintained at 4° C. for 2 weeks. But he implies that these changes may be merely changes in solubility rather than in actual content.

That adenine can be converted to

guanine *in vivo*, probably at the nucleotide level (phosphate, sugar, and a single base), is suggested by results in which rats utilized adenine containing heavy nitrogen ( $N^{15}$ ), and the nitrogen was later found in the adenine and guanine of tissue nucleic acids (White et al., 1959). They maintained that the incorporation of adenine into adult rat tissue is mainly in RNA. Incorporation of adenine into DNA is usually slow or negligible; however, in partially hepatectomized animals undergoing rapid liver regeneration, there is extensive incorporation of adenine into DNA. On the basis of these and similar results, White et al. suggested that DNA is formed to an appreciable extent only during active mitosis by a cell.

White, Handler, Smith, and Stattan (1959) also reported that the cytosine nucleotide can be obtained from chemical reactions involving the uracil nucleotide. In rats a radioactive isotope labeled cytosine compound was a precursor of both cytosine and uracil.

Research with purine and pyrimidine analogues, which are basically similar to the natural bases but differ in one or more side bonds, has provided some interesting results. Dunn and Smith (1954) have shown that in *Escherichia coli*, 5-bromouracil is incorporated into DNA and replaces thymine. They found that the addition of either 5-bromouracil or 5-iodouracil inhibited the natural reactions of *E. coli*. The mean generation time of the bacteria was increased 2 to 4 hours after addition of the base analogues; the degree of inhibition increased with greater amounts of base analogues. However, inhibition could be reversed by thymine. After inhibition by either analogue, the proportion of bacteria giving colonies on nutrient agar was reduced. When all of the thymine was replaced by bromouracil, only 9% of the phage particles yielded viable progeny (Strauss, 1960).

The base analogues have been used frequently in the DNA of bacteriophages (Sinsheimer, 1960). For example, 5-bromouracil and 2-aminopurine are presumed to act to bring about the replacement of the adenine-thymine pairs by guanine-cytosine ones, and vice versa. These agents also bring about a reversion of these changes. Freese (1961) has indicated that low pH, ethyl ethane sulfonate, and other agents, will cause transitions from one pair to the other.

Benzer (1961, 1962) has made a detailed examination of a small portion of the genetic map of bacteriophage T4, a portion which controls the ability of the phage to grow in *E. coli*. He indicated that A-T pairs are held much less strongly than are G-C pairs which suggests that in mutation the A-T pair "hot spots" will be more subject to substitution. He stated that A-T pairing would change to A-G and thereon to C-G.

In a few organisms, for example, tobacco mosaic virus, polio virus, and influenza virus, no DNA is present. In this case RNA appears to be the hereditary material. Nitrous acid has been used as a mutagen with tobacco mosaic virus RNA. This reagent substitutes hydroxyl groups (OH) for amine groups (NH<sub>2</sub>). Strauss (1960) has indicated that nitrous acid reacts with nucleic acids containing adenine, guanine, and cytosine and converts them to the corresponding base analogues containing hydroxyl compounds, hypoxanthine, xanthine, and uracil. These results imply that the reaction of nitrous acid with nucleic acids produces base analogues which result in mutation upon the duplication of genetic material. Tsugita and Fraenkel-Conrat (1960) have shown that nitrous acid alters the composition of RNA of tobacco mosaic virus and that the resulting protein of the mutant differed from the parent strain with three amino acids being replaced by three

others (proline, aspartic acid, and threonine by leucine, alanine, and serine).

These changes have involved the use of chemical agents to bring about a change of base composition. But there is no information to indicate that modification in the DNA molecule can occur with moderate stimulation except for the White et al. (1959) material, discussed above. In fact biochemists are skeptical about such changes because of the DNA stability. However, the DNA molecule should not be eliminated completely for several reasons. First, most of the DNA which has been investigated has been from nonneural tissue. It is possible that the activity of DNA in nerve cells is different than its activity in other cells and may be more conducive to modification. Mirsky (1953) indicated that the metabolic activity of protein varies greatly in different nuclei of an organism, suggesting that the chromosomes of different cells vary in their activity. He showed that differential activity did occur. With heavy nitrogen in glycine the DNA in mouse liver chromosomes took up nitrogen three times as fast as that in the kidney, even though the liver and kidney have equal amounts of DNA. Also the cells of a digestive gland such as the pancreas are far more active when an animal is fed. While these cells are active, the DNA of their chromosomes takes up 50% more labeled nitrogen. Possibly the DNA of nerve cells is more active than that in most other tissue.

Furthermore, the stability of the DNA molecule does not necessarily preclude it operating as a source for memory changes. In fact its stability would appear to be an argument in its favor. The experiential code requires great stability. If memory were maintained by molecules which were too active metabolically, then it would change rapidly causing chaotic behavior. In any event, it may be that changes at the

base sites do occur. Yet such an interpretation must be held with caution.

### RNA

Information concerning the possibility of RNA playing a role in memory is more encouraging at first sight. RNA varies from cell to cell and is very active metabolically (Ris, 1957). Sinsheimer (1960) reported that the amount of RNA in the salivary gland of *Drosophila* drops rapidly during early stages of differentiation whereas the amount of DNA increases, and changes in the overall nucleotide composition of RNA of *Chlorella* cells occur during starvation. LeBaron (1959) indicated that other cytochemical research provides evidence for increased activity of cellular RNA and proteins. He concluded that:

There is certainly ample evidence for the active turnover of various lipide, protein and nucleic acid structural constituents, and the possibility exists that there is some alteration of this turnover on stimulation [p. 897].

Hydèn (1961) reported that the RNA content of the nerve cell ranks with the highest of all cells in the body. He showed that in man the RNA content of the motor nerve cells in the spinal cord increases from the third year of life to age 40, remains constant to about 60, and then declines rapidly thereafter. He found that if an animal is deprived of stimulation in one of the sensory systems, for example, in vision or hearing, the neurons in that system do not develop biochemically. The structure appeared normal but the nerve cell was impoverished in both RNA and proteins. Hydèn maintained that individuals with certain psychic disorders have smaller amounts of RNA and proteins in ganglion cells of the central nervous system than do normal individuals (cited by Davidson, 1960). Administration of malononitrile to these individuals increased the content of these substances. Egyhazi and Hydèn (1961) indicated

that the malononitrile action is due to the formation of a dimer of malononitrile, tricyanoamino-propane. They reported that small amounts of this latter compound caused a remarkable increase in the amounts of proteins and RNA in the cell and modified the base composition of the RNA with guanine increasing by almost 300%.

Riesen (1958), in discussing his work on light deprivation, referred to Brattgard's findings that the content of RNA and proteins decreases in retinal ganglion cells with prolonged light deprivation. He thought that RNA and protein were so highly susceptible to recent prior stimulation as to obviate their being considered as a mechanism for durable memory. Instead he reasoned that they might be important for immediate memory. Pertinent to this thought is the work of Geiger, Yamasaki, and Lyons (1958). These individuals stimulated the brain cortex of cats and found a change of RNA in the stimulated areas, which was reversible in minutes.

Likewise, Morrell (1961) has shown gross changes in RNA of nerve cells in a cellular learning-like situation. He stimulated a portion of the cortex with ethyl chloride spray. The homologous tissue of the opposite hemisphere showed activity during this stimulation. At first the activity in the opposite hemisphere appeared only when the stimulated cortex was active. After a time, the tissue in the nonstimulated area showed spontaneous excitation even when it was isolated from the stimulated hemisphere by cutting its connections. Biochemical analysis of the neurons in the isolated tissue showed a change in the RNA content. Kreps has also reported an altered RNA turnover in conditioning (cited by Gerard, 1960).

The above results have been pertinent to sensory stimulation experiments. However, several experiments have been concerned with memory functions. In-

direct support for the RNA modification hypothesis is provided by the results of Cameron and Solyom (1961). They found that administration of RNA (but not DNA) to aged individuals brought about memory improvement. These changes involved almost total retention in some cases. Later, when the RNA was discontinued memory relapses occurred. However, it is possible that the RNA was merely supplying a nutritional need and would improve any type of performance. This result should be considered relative to Hyden's finding that the amount of RNA decreases after the age of 60.

An experiment by Corning and John (1961) is also pertinent. Using a classical conditioning procedure, pairing light with shock, they conditioned a number of flatworms and then transected them into head and tail sections. Previous experimentation had indicated that heads will regenerate new tails, tails will regenerate new heads, and both will retain some "memory" of the avoidance situation. Corning and John thought that RNA might play a role in the transmission of an acquired structural configuration from the trained to the regenerating tissues. Thus they reasoned that if the trained portions were regenerated in the presence of RNase (ribonuclease), the enzyme would affect the altered RNA structure, producing some animals with a naive head (regenerated portion) and trained tails and others with trained heads and naive tails (regenerate). They stated that the head region would probably be dominant and thus the trained head animals should show more retention. Their results indicated that heads regenerated in RNase retained the memory as well as did head and tail sections regenerating in pond water. However, the tails regenerating in RNase performed randomly. The authors suggested that the RNase did not affect

intact tissue but did interfere with regenerating tissue and maintained that the results are compatible with the assumption that RNA is a memory mechanism.

The idea that changes in the linear sequence of bases in RNA constitutes the experiential code has been offered independently by a number of individuals. However, Dingman and Sporn (1961) suggested that changes in the helical structure and overall configuration, as well as sequence changes, could be the basis for memory. They performed two experiments with radioactive 8-azaguanine injections in rats which were pertinent to the RNA hypothesis. 8-azaguanine was used as an inhibitor of RNA because this base analogue had been shown to be an inhibitor of enzyme synthesis in bacteria. In both experiments paper chromatographic procedures indicated that the base analogue had been incorporated into the RNA of the brain. In neither experiment was there any significant difference between experimental and control animals in average time to run the maze, suggesting that 8-azaguanine had no adverse effect on the motor ability of the animals. However, in one study the experimental animals had a greater average number of errors than did the controls on all 15 trials in the learning of a maze. In another experiment concerned with retention of a maze pattern (tested by a single trail after learning a maze), experimental animals did not differ significantly from control animals even though the experimental animals had a greater average number of errors than did the control animals. There were only 8 animals used in each group (as compared with 14 and 15 in the learning experiment); thus it is possible that if  $n$  had been larger in the retention experiment, the results would have indicated that 8-azaguanine adversely affects

both learning and retention of maze patterns in rats.

Based on their results Dingman and Sporn maintained that RNA may be directly involved in learning but not in retention. However, they admitted that their results did not necessarily indicate that RNA metabolism was intimately linked with the formation of memory traces in the brain because 8-azaguanine might have interfered with metabolic processes which affected RNA indirectly.

Most of the above studies in which RNA changes were reported do not exclude the possibility that the basic changes were effected in DNA which then brought about changes in RNA. However, the results of Cameron and Solyom and of Corning and John appear to argue against the involvement of DNA in memory. But both experiments appear to suffer from methodological defects. The work by Cameron and Solyom lacked the use of doubleblind or similar techniques to eliminate experimenter bias in all experiments.

The results of Corning and John are not as clear-cut as they appear. They performed multiple *t* tests (exact number not specified) and reported only the significant ones. It is possible that some of these were Type I errors in that with many groups such as Corning and John had, the probability of Type I errors is greater than that specified in *t* tables. Furthermore, from a qualitative analysis of their basic table of data there is a suggestion that RNase may have a residual effect, sensitizing or irritating the animals such that they are predisposed to make a great number of responses to light. In the retention data three groups which were in RNase (other than tail animals) showed a greater number of CRs to light on each of 3 days than did corresponding control groups in pond water. These results would probably be significant if a re-

peated measurements analysis were attempted. Because of these considerations the Corning and John study is inconclusive.

The big problem faced by advocates of the RNA hypothesis is that RNA is not of a single homogeneous type. There are several types of RNA which vary in structure and function. One type, messenger RNA, which receives the basic genetic code from a strand of DNA (Hall & Spiegelman, 1961; Hurwitz & Furth, 1962) would appear to be the most important one to serve as an experiential code. Another type, soluble RNA, which functions to carry activated amino acids to the ribosomes for protein synthesis, and other RNA in the ribosomes might be considered as "junk" for information coding purposes. Thus, the changes in amounts of RNA, or base changes, demonstrated by Hyden and others may be relative to junk RNA through metabolic activity and be totally unrelated to information coding. The reversible changes in RNA content reported by Geiger et al. (1958) and by Riesen (1958) appear to support this possibility.

Therefore, even though there is more optimism among scientists about the RNA hypothesis than the DNA hypothesis, there is no conclusive (or even good suggestive) evidence that the former is more appropriate than the latter. Furthermore, even though the experimental results superficially tend to suggest the involvement of RNA in memory, its extreme metabolic reactivity raises some doubt as to how stable memory can be handled by an overly reactive molecule. On the other hand, the extreme reactivity may reflect gross aspects, with base changes responsible for stable memory less involved in overall reactivity. But, as has been indicated above, Riesen's suggestion that RNA subserves an immediate memory function may be entirely appro-

priate and require that another molecular mechanism be postulated for maintaining permanent memory.

### Conclusions

If we consider memory as of two types, transient and permanent (Hebb, 1949; Jeffress, 1951), and use DNA and RNA as candidates for these functions, we have four possibilities as indicated in Table 1. However, we should add two possibilities which correspond to DNA or RNA serving only for permanent memory whereas a short-term phenomenon such as reverberatory circuits could suffice for transient memory. We should also add two other possibilities which involve DNA or RNA functioning for transient memory with reverberatory circuits maintaining permanent memory.

Of the eight possibilities, A and C (which use DNA for transient memory) seem to be negated because of the relatively great stability of DNA. The Cameron and Solyom (1961) and Corning and John (1961) results might suggest eliminating Possibility D. Evidence (see below) suggesting that reverberation ceases within an hour would preclude Possibilities G and H. Intuitively the possibilities in which reverberatory circuits maintain short-term memory until changes can be effected in DNA or RNA (E and F) seem most likely.

Morrell (1961, p. 478) cites data by Chow which is consistent with this possibility. Chow reported that electrical

changes in neural tissue were detected only during the early stages of conditioning. Morrell uses these results, as well as others indicating that memory survives convulsions, electroshock, trauma, concussion, and deep anesthesia, to implicate RNA as a possible durable memory molecule. Likewise, John (1961, p. 465) stated that reverberatory circuits could be the basis of short-term memory but that long-term memory would seem to necessitate a structural modification.

That reverberatory circuits are the basis for short-term memory appears plausible because of much experimental data. For example, experiments have shown that individuals who learn something and then go to sleep tend to retain more than do individuals who do not sleep after learning (Crafts, Schneirla, Robinson, & Gilbert, 1950). Disruption of the "consolidation process" by shocking animals immediately after a learning event (running a maze) is reported to interfere with the memory for the event (Duncan, 1949). Shocks introduced less than a minute after the run interfered more than did shocks given later. Gerard (1953) found similar results with hamsters. These results appear to indicate that some transient activity is responsible for memory, and if other activities intervene there is less chance for the permanent changes to occur. Likewise, observations with elderly individuals support the notion of reverberatory circuits. In these people

TABLE 1

SOME PLAUSIBLE NEUROCHEMICAL POSSIBILITIES IN DEALING WITH TRANSIENT AND PERMANENT MEMORY

	Possibilities							
	A	B	C	D	E	F	G	H
Transient	DNA	RNA	DNA	RNA	Reverberatory Circuits		DNA	RNA
Permanent	DNA	RNA	RNA	DNA	DNA	RNA	Reverberatory Circuits	

we know that the brain tissue is less plastic and hardening of blood vessels and other tissues occurs. If the tissue is less plastic, then it would seem that the initial reverberatory circuits would be less effective in developing neural changes. Thus the retention of the learning of new materials would be less than when the individual were younger. This is exactly what one finds with elderly persons, with difficulty in the learning of new materials and quick forgetting of newly learned material.

Other individuals have also suggested a dual process for memory. For example, Dethier and Stellar (1961) maintained that brain-damaged studies with man and animal-learning experiments involving shock suggest that memory is a two-part process: an initial, vulnerable process lasting 15 minutes to an hour; and a later, invulnerable process providing the permanent basis for memory.

Glickman (1961) suggested essentially the same, a reverberatory circuit hypothesis for consolidation of memory traces. He cited work by Burns (1954, 1958) which is consistent with the hypothesis. Burns isolated small areas of the cortex and noted their responses to electrical stimulation. He found that single trains of pulses can initiate activity lasting for 30 minutes or more, that later electrical stimulation interferes with this activity, and that the activity becomes easier to evoke with repetitions of the stimulus.

The two types of memory would depend upon different neurological sites (Gaito, 1961). The reverberatory circuits would depend upon synaptic transmission and be necessary (but not sufficient) to effect changes in DNA or RNA elsewhere in the cell. The changes in the nucleic acids would appear to occur in the soma of brain nerve cells in that DNA is located in the nucleus and Nissl substance, which contains RNA, is more

profuse in the cytoplasm of the cell body than in either the dendrites or in the axons (Hydén, 1959).

In conclusion, even though much microbiological evidence indicates that DNA is the genetic code, microbiologists are skeptical about its possibility as the experiential code. Yet it has not clearly been eliminated. RNA, however, appears superficially to be implicated in memory events and may be the experiential code. The experimental results are not definitive as yet. RNA may be the basic molecule for the experiential code or it may be indirectly involved through the effect of a more basic molecule or enzyme. If RNA is the experiential code, the exact mechanism of coding is uncertain. The linear sequence of purines and pyrimidines appears to be the most plausible mechanism in that these sequences provide the basis for genetic coding in DNA. However, the suggestions of Dingman and Sporn (1961) that the helical structure and overall configuration are other possible mechanisms should be seriously considered.

Thus, there is definite evidence to indicate that gross changes in DNA and RNA can occur during stimulation. However, no one has directly detected a change of submolecular structure in either DNA or RNA such as discussed above; all the evidence for changes are of indirect nature. The hypothesizing about nucleic acids must remain in the realm of speculation. The validity of these hypotheses must await further research by biochemists, psychologists, neurophysiologists, and others of related areas.

#### OVERALL EVALUATION

A neurochemical approach to learning should attempt to describe the biochemical events which underlie the major learning phenomena, viz., mechanisms for short-term memory, long-term

memory, reactivation, normal forgetting, special forgetting (such as in hypnosis, repression, amnesia), and overall integration. Let us look at the neurochemical approaches relative to each of these phenomena.

### *Short- and Long-Term Memory*

At present there are essentially two approaches to the representation of experience, i.e., memory. The first stems from the Hebbian influence and involves the linking together of neural units which by their linkage symbolize experiences. The molar group employ this conceptualization. Short-term memory is maintained by reverberation through various circuits until some synaptic change can occur which will then increase the probability of excitation throughout the circuitry when later some portions are excited. The synaptic change for Overton is calcium displacement. The California group do not concern themselves with short- and long-term memory but the synaptic events which they stress are the ACh-ChE activities.

The second means of representing experience is symbolically by a chemical code, viz., the linear sequence of bases in DNA or RNA as hypothesized by the molecular approach.

Actually both approaches could allow for reverberatory circuits to maintain transient memory. Differences, however, would appear in handling permanent memory. The molecular group would allow for changes to occur at the synapse but these changes would not be the end result. They would merely be preliminary to a change of the code elsewhere in the cell; i.e., they would be necessary, but not sufficient, for learning.

### *Reactivation*

The California group are not concerned with this or the remaining events

we will consider. Overton allows the synaptic changes (calcium displacement) to facilitate the reactivation of the reverberatory circuits to bring about recall. Thus he treats permanent memory and reactivation as if they were the same. However, it appears that some distinction should be made between the "trace" and the process which reactivates the trace (Koffka, 1935).

Of the molecular group only Hyden has attempted to explain a reactivation process. A change in one or more bases in RNA is effected through frequency modulation. The changed RNA will synthesize new protein which differs in amino acid sequence from protein synthesized by the previous RNA. When later the nerve cells are stimulated by the same frequency the protein is activated, dissociates, and then the dissociated products combine with a complementary molecule. These events activate a transmitter substance (ACh or an inhibitory substance) and the post-synaptic structure is excited. Thus reactivation requires that the modified protein be excited by nerve impulses which have the same frequency as those which previously modified the RNA.

### *Normal and Special Forgetting*

Only Overton has considered these events. Normal forgetting results from disruption of reverberatory circuits, presumably by increased calcium deposition. To handle special forgetting, viz., repression, he maintained that some reverberatory circuits are protected by the jamming of other neural activity which might disrupt those circuits. Jamming represents the repression of immediate or recalled memories. However, Overton did not indicate the neurological or biochemical basis for jamming.

### *Integration*

An attempt to describe overall integration biochemically would appear to

require adequate coverage of the above events. Thus it is understandable that none of the approaches have talked of integration. However, Overton heads somewhat in that direction when he describes insight as the combining of two or more reverberating circuits into one large circuit (*a la* Hebb).

This section indicates that the neurochemical approaches are incomplete. This is to be expected inasmuch as these approaches are relatively new. The two approaches need not be considered as antagonistic; they appear to be of complementary nature in that the description of each is on a different level. The ACh-ChE activity, and possible acetic acid etching of calcium compounds, may be the important events which occur at the synapse and facilitate the modification of DNA and RNA elsewhere in the cell, or vice versa. One could even add the neurological approach of Hebb (1949) with cell assemblies and phase sequences and have further complementarity. The three approaches describe events on different levels of discourse and can be valuable supplements of each other.

Of the various proponents of the neurochemical approach, the California group sticks closest to empirical results and thus resorts to little speculation. The others use empirical data as a springboard for their speculations. Obviously such speculations may be completely wrong in the ultimate analysis; however, they provide investigators with a number of provocative ideas to guide experimental research.

The specificity of the molecular approach is a definite advantage over the molar approach in that a description of events at submolecular levels may provide biological scientists with the "atoms" of behavior. Furthermore, behavioral scientists have always been interested in the relationship between heredity and environment. The molec-

ular advocates are attempting to handle this relationship at a molecular level. Gaito would have the two interacting within the same molecule (DNA) or through two molecules. In the latter case DNA provides the basic code to be passed on to messenger RNA, which is modified through experiential events. This interpretation would also be the viewpoint of Dingman and Sporn and of Hyden.

The authors believe that the molecular viewpoint will be the most useful within the next decade or so in that rapid advances in microbiology will force behavioral scientists to include the nucleic acids in their schema. However, the ultimate value of both approaches will be determined by experimentation.

#### REFERENCES

- ALFERT, M. Some cytochemical contributions to genetic chemistry. In W. P. McElroy & R. Glass (Eds.), *The chemical basis of heredity*. Baltimore: Johns Hopkins Press, 1957.
- BENNETT, E. L., KRECH, D., ROSENZWEIG, M. R., KARLSSON, H., DYE, N., & OHLANDER, A. Cholinesterase and lactic dehydrogenase activity in the rat brain. *J. Neurochem.*, 1958, 3, 153-160.
- BENNETT, E. L., ROSENZWEIG, M. R., KRECH, D., KARLSSON, H., DYE, N., & OHLANDER, A. Individual, strain and age differences in cholinesterase activity of the rat brain. *J. Neurochem.*, 1958, 3, 144-152.
- BENZER, S. On the topography of the genetic fine structure. *Proc. Acad. Nat. Sci.*, 1961, 47, 403-415.
- BENZER, S. The fine structure of the gene. *Scient. American*, 1962, 206, 70-84.
- BROADBENT, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.*, 1957, 64, 205-215.
- BURNS, B. D. The promotion of afterbursts in isolated unanesthetized cerebral cortex. *J. Physiol.*, 1954, 125, 427-446.
- BURNS, B. D. *The mammalian cerebral cortex*. London, England: Arnold, 1958.
- BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.
- BUTLER, J. A. V. *Inside the living cell*. New York: Basic Books, 1959.
- CAMERON, D. E., & SOLYOM, L. Effects of

- ribonucleic acid on memory. *Geriatrics*, 1961, 16, 74-81.
- CHARGAFF, E. First steps toward a chemistry of heredity. In O. Hoffman-Ostenof (Ed.), *Fourth international congress of biochemistry*. Vol. 14. *Transactions of the plenary sessions*. New York: Pergamon Press, 1958.
- CORNING, W. C., & JOHN, E. R. Effect of ribonuclease on retention of conditioned response in regenerated planarians. *Science*, 1961, 134, 1363-1365.
- CRAFTS, L. W., SCHNEIRLA, T. C., ROBINSON, E. E., & GILBERT, R. W. *Recent experiments in psychology*. New York: McGraw-Hill, 1950.
- CRICK, F. H. C. The structure of the hereditary material. *Scient. American*, 1954, 191, 54-61.
- CRICK, F. H. C. Nucleic acids. *Scient. American*, 1957, 197, 188-200.
- DAVIDSON, J. H. *The biochemistry of the nucleic acids*. New York: Wiley, 1960.
- DETHIER, V. G., & STELLAR, E. *Animal behavior*. Englewood Cliffs, N. J.: Prentice-Hall, 1961.
- DINGMAN, W., & SPORN, M. B. The incorporation of 8-azaguanine into rat brain RNA and its effect on maze-learning by the rat: An inquiry into the biochemical bases of memory. *J. psychiat. Res.*, 1961, 1, 1-11.
- DUNCAN, C. P. Retroactive effect of electroshock on learning. *J. comp. physiol. Psychol.*, 1949, 42, 32-44.
- DUNN, D. B., & SMITH, J. D. Incorporation of halogenated pyrimidines into the deoxyribonucleic acids of *Bacterium coli* and its bacteriophages. *Nature*, 1954, 174, 305.
- EGYHAZI, E., & HYDÉN, H. Experimentally induced changes in the base composition of the ribonucleic acids of isolated nerve cells and their oligodendroglial cells. *J. Biophys. Biochem. Cytol.*, 1961, 10, 403-410.
- ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 94-107.
- FRESE, A. B. Transitions and transversions induced by depurinating agents. *Proc. Nat. Acad. Sci.*, 1961, 47, 540-545.
- GAITO, J. A biochemical approach to learning and memory. *Psychol. Rev.*, 1961, 68, 283-292.
- GEIGER, A., YAMASOKI, S., & LYONS, R. Changes in nitrogenous components of brain produced by stimulation of short duration. *Amer. J. Physiol.*, 1958, 184, 239-243.
- GERARD, R. W. What is memory? *Scient. American*, 1953, 189, 118-126.
- GERARD, R. W. Neurophysiology: An integration (molecules, neurons, and behavior). In J. Field (Ed.), *Handbook of physiology-neurophysiology*. Vol. 3. Baltimore: Williams & Wilkins, 1960.
- GLICKMAN, S. E. Perseverative neural processes and consolidation of the memory trace. *Psychol. Bull.*, 1961, 58, 218-233.
- HALL, B. J., & SPIEGELMAN, S. Sequence complementarity of T2-DNA and T2-specific RNA. *Proc. Nat. Acad. Sci.*, 1961, 47, 137-146.
- HALSTEAD, W. C. Brain and intelligence. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951.
- HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
- HURWITZ, J., & FURTH, J. J. Messenger RNA. *Scient. American*, 1962, 206, 41-49.
- HYDÉN, H. Biochemical changes in glial cells and nerve cells at varying activity. In *Proceedings of the fourth international congress of biochemistry: Biochemistry of the central nervous system*. Vol. 3. London: Pergamon Press, 1959.
- HYDÉN, H. Satellite cells in the nervous system. *Scient. American*, 1961, 205, 62-70.
- JEFFRESS, L. A. (Ed.) *Cerebral mechanisms in behavior*. New York: Wiley, 1951.
- JOHN, E. R. High nervous functions: Brain functions and learning. In V. E. Hall, F. A. Fuhrman, & A. C. Giese (Eds.), *Annual review of physiology*. Palo Alto, Calif.: Annual Reviews, 1961.
- KATZ, J. J., & HALSTEAD, W. C. Protein organization and mental functions. *Comp. Psychol. Monogr.*, 1950, 20(1, Ser. No. 103).
- KOFFKA, R. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
- KORNBERG, A. The biological synthesis of deoxyribonucleic acid. Paper read at Karolinska Institute, Stockholm, Sweden, 11 December 1959.
- KRECH, D., ROSENZWEIG, M. R., & BENNETT, E. L. Dimensions of discrimination and level of cholinesterase activity in the cerebral cortex of the rat. *J. comp. physiol. Psychol.*, 1956, 49, 261-268.
- KRECH, D., ROSENZWEIG, M. R., & BENNETT, E. L. Correlation between brain cholinesterase and brain weight within two strains of rats. *Amer. J. Physiol.*, 1959, 196, 31-32.
- KRECH, D., ROSENZWEIG, M. R., BENNETT, E. L., & KRUECKEL, B. Enzyme concentration in the brain and adaptive behavior patterns. *Science*, 1954, 120, 994-996.
- KRECH, D., ROSENZWEIG, M. R., BENNETT, E. L., & LONGUEIL, C. L. Changes in brain chemistry of the rat following experience. Paper read at American Psychological Association, Cincinnati, 1959.
- LASHLEY, K. S. *Brain mechanisms and intelligence*.

- gence. Chicago: Univer. Chicago Press, 1929.
- LEBARON, F. W. Neurochemistry. In J. W. Luck, F. W. Allen, & G. Mackinney (Eds.), *Annual review of biochemistry*. Palo Alto, Calif.: Annual Reviews, 1959.
- LEDERBERG, J. A view of genetics. *Science*, 1960, 131, 260-273.
- MILLER, G. A. What is information measurement? *Amer. Psychologist*, 1953, 8, 3-11.
- MILNER, P. M. The cell assembly: Mark II. *Psychol. Rev.*, 1957, 64, 242-252.
- MIRSKY, A. B. The chemistry of heredity. *Scient. American*, 1953, 188, 47-57.
- MORRELL, F. Electrophysiological contributions to the neural basis of learning. *Physiol. Rev.*, 1961, 41, 442-494.
- OVERALL, J. E. A cognitive probability model for learning. *Psychometrika*, 1960, 25, 159-172.
- OVERTON, R. K. An effect of high- and low-calcium diet on the maze performance of rats. *J. comp. physiol. Psychol.*, 1958, 51, 697-699.
- OVERTON, R. K. The calcium displacement hypothesis: A review. *Psychol. Rep.*, 1959, 5, 721-724. (a)
- OVERTON, R. K. *Thought and action: A physiological approach*. New York: Random House, 1959. (b)
- PAGE, J. H. Thudichum and the chemistry of the brain. In K. A. C. Elliot, I. H. Page, & J. H. Quastel (Eds.), *Neurochemistry*. Springfield, Ill.: Charles C Thomas, 1955.
- PAULING, L. *The nature of the chemical bond*. Ithaca, N. Y.: Cornell Univer. Press, 1960.
- PRIBRAM, K. H. A review of theory in physiological psychology. In P. R. Farnsworth & Q. McNemar (Eds.), *Annual review of psychology*. Palo Alto, Calif.: Annual Reviews, 1960.
- RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, 62, 11-19.
- RIESEN, A. H. Plasticity of behavior: Psychological aspects. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958.
- RINKEL, M., & DENBER, H. C. B. (Eds.) *Chemical concepts of psychosis*. New York: McDowell, Obolensky, 1958.
- RIS, H. Chromosome structure. In W. D. McElroy & B. Glass (Eds.), *The chemical basis of heredity*. Baltimore: Johns Hopkins Press, 1957.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.*, 1958, 65, 386-408.
- ROSENZWEIG, M. R. Effects of heredity and environment on brain chemistry, brain anatomy, and learning ability in the rat. Paper presented at Kansas symposium on physiological determinates of behavior: Implications for mental retardation, University of Kansas, 4 June 1963.
- ROSENZWEIG, M. R., KRECH, D., & BENNETT, E. L. Effects of pentobarbital sodium on adaptive behavior patterns in the rat. *Science*, 1956, 123, 371-372.
- ROSENZWEIG, M. R., KRECH, D., & BENNETT, E. L. A search for relations between brain chemistry and behavior. *Psychol. Bull.*, 1960, 57, 476-492.
- RUBIN, L. S. Recent advances in the chemistry of psychotic disorders. *Psychol. Bull.*, 1959, 56, 375-383.
- SAGER, R., & RYAN, F. J. *Cell heredity*. New York: Wiley, 1961.
- SINSHEIMER, R. L. The biochemistry of genetic factors. In J. M. Luck, F. W. Allen, & G. Mackinney (Eds.), *Annual review of biochemistry*. Palo Alto, Calif.: Annual Reviews, 1960.
- STANLEY, W. M., & VALENS, E. G. *Viruses and the nature of life*. New York: E. P. Dutton, 1961.
- STRAUSS, B. S. *An outline of chemical genetics*. Philadelphia, Pa.: Saunders, 1960.
- THOMAS, G. J. Neurophysiology of learning. In P. R. Farnsworth, O. McNemar, & Q. McNemar (Eds.), *Annual review of psychology*. Palo Alto, Calif.: Annual Reviews, 1962.
- THUDICHUM, J. L. W. *A treatise on the chemical constitution of the brain*. London: Bailliere, Tindall, & Cox, 1884.
- TOWER, D. B. The neurochemical substrates of cerebral function and activity. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958.
- TSUGITA, A., & FRAENKEL-CONRAT, H. The amino acid composition and C-terminal sequence of a chemically evoked mutant of TMV. *Proc. Nat. Acad. Sci.*, 1960, 46, 636-642.
- WECHSLER, D. *The measurement of intelligence*. Baltimore, Md.: Williams & Wilkins, 1944.
- WIENER, N. *Cybernetics*. New York: Wiley, 1948.
- WILLIAMS, R. J. *Biochemical individuality*. New York: Wiley, 1956.
- WHITE, A., HANDLER, P., SMITH, E. L., & STATTAN, D. *Principles of biochemistry*. New York: McGraw-Hill, 1959.

(Received July 5, 1962)

## VERBAL LEARNING IN CHILDREN

GEOFFREY KEPPEL<sup>1</sup>

Northwestern University

The major purpose of the review was to examine the validity of the assumption that the laws of verbal learning obtained on the college student will hold reasonably well for younger populations of Ss. Various areas of verbal learning, such as free learning, paired-associate learning, serial learning, transfer, retention, and retroactive-proactive inhibition, were reviewed in order to determine the extent to which the general findings of these areas, based on experiments in which the college student has served as S, have been investigated and replicated with children as Ss. The literature on children's verbal learning revealed a reasonable correspondence in the effective variables reported and in the relationships that have been identified with adult Ss.

For the past 2 decades subjects in experiments on verbal learning have been preponderantly the college sophomore, with only an occasional experiment reported in which younger or older subjects were employed. One obvious reason for the heavy use of college students is their ready availability. But a second reason may be an implicit assumption that the laws of verbal learning obtained on the college student will hold reasonably well for other populations. The major purpose of the present review is to examine the validity of this assumption. This will be done by comparing the results of the available studies in verbal learning which have used children with the results obtained in experiments where the college student has been used as the subject.

Comparisons of the above nature may be used to answer several questions. First, we may ask whether the various phenomena of verbal learning derived from the college student have in fact been investigated with subjects of younger ages. Second, if such studies have been done, we may ask if the same *general* phenomena emerge with both children and college students; for example, is retroactive inhibition produced at various age levels? Third, we may

ask if the *specific* laws relating a phenomenon to various manipulable variables are the same for both college students and children; for example, is the relationship between amount of retroactive inhibition and degree of interpolated learning the same? Obviously the ability to answer these questions—questions which increase in specificity—depends upon the availability of appropriate data. It will be seen that most of the review must necessarily be confined to answering the first two questions, namely, are studies available, and, if so, how do the *general* findings compare.

Certain restrictions were placed on the studies to be reviewed.

1. From the studies using the college student as subject, only phenomena which are judged to be firmly established experimentally will be considered. For example, reminiscence cannot be consistently produced with college students, and it would be of little value for the present review to ask whether or not it can be produced with children.

2. For any given reliable phenomenon derived from experiments using college students, no attempt will be made to cite complete references. Rather, illustrative references will be given.

3. Exclusion of studies of verbal or language behavior.

<sup>1</sup> Now at University of California, Berkeley.

4. Exclusion of studies using non-verbal responses.

5. The starting date for this review is 1940, with a selection being made of earlier studies. Reviews are available of studies of verbal learning in children conducted prior to 1940 (e.g., McGeoch, 1942; McGeoch & Irion, 1952; Munn, 1954; Peterson, 1933; Spiker, 1960b).

The review will follow an arbitrary outline which first considers the simpler experimental operations in verbal learning, followed by the more complex. Although not considered systematically, several studies concerned with the associative or verbal repertoire of the subject will be discussed first, followed by various methods by which verbal learning is studied, specifically, free learning, paired-associate learning, serial learning, and other learning methods. A second general classification pertains to the more classical topics of verbal learning in which two-stage effects are studied, such as transfer, simple or single-list retention, and retroactive and proactive inhibition. Finally, the implications of children learning to the theories of verbal learning will be considered. Throughout the review, the term "children" will be used to refer to all subjects who have not reached college age.

#### ASSOCIATIVE REPERTOIRE

The learning that is observed in any verbal-learning experiment is superimposed upon all language learning that has occurred prior to the laboratory experiment; it is this network of language habits that may be termed the subject's verbal or associative repertoire. This repertoire may consist of various components, such as the subject's repertoire of words and letters, the associations between these words (e.g., Russell & Jenkins, 1954) and between letters (e.g., Underwood & Schulz, 1960), and any learned habits of combinations of letters or words, e.g., syntactical-gram-

matical habits. The free-association technique has provided a useful method of assessing the associative repertoire of adults and children; the older studies comparing free-association norms of children and adults are reviewed by Woodworth and Schlosberg (1954, pp. 52-56). Recently, Brown and Berko (1960) compared the free associations of first-, second-, and third-grade children and adults with regard to the tendency to give responses from the same form class as the stimulus word; for all ages, nouns and adjectives produced more such homogeneous responses than other parts of speech and the tendency to give these responses was found to be a direct function of age. Similarly, Ervin (1961) found a significant increase with age (Grades 1-6) in the proportion of responses from the same grammatical class as the stimulus word.

Letter-association norms, obtained for college students (Underwood & Schulz, 1960, Ch. 9), have been useful in specifying extraexperimental sources of letter-to-letter interference for two- and three-letter units (e.g., Underwood & Postman, 1960). In an unpublished study, the author obtained single-letter responses to a sample of single- and double-letter stimuli from high school students and from second-grade students. The frequency of single-letter responses to the single- and double-letter stimuli for the two groups was highly correlated with the frequency shown by Underwood and Schulz, product-moment correlations of approximately .80 and .60 being obtained for high school and second-grade students, respectively. The magnitude of these correlations implies a relatively early learning of letter-language habits that appear to be of some consequence to the adult verbal learner. Furthermore, the relatively high correspondence in letter associations for the three groups may provide for differential

predictions of the learning rates as a function of meaningfulness for various age groups. For example, it should be possible to construct lists of materials varying in meaningfulness but equal in difficulty at each meaningfulness level for the various age groups as inferred from the respective letter-association norms. Assuming that the strength of the letter-language habits is a direct function of the age of the subject, then an interaction between meaningfulness and age may be obtained, variations in meaningfulness producing smaller differences in learning rates for children than for college students. Specifically, this prediction follows if it is assumed that the letter habits of college students tend to produce more interference in the learning of low meaningful material and more facilitation in the learning of high meaningful material than for the child whose letter-language habits are not as strong.

These developmental studies of word and letter associations may provide the verbal-learning experimenter with an understanding of the various determinants of the adult free-association norms and an appreciation of the complex relationships existing among verbal units.

#### FREE LEARNING

Perhaps the simplest learning situation, free learning usually involves the single presentation of a series of words and the subsequent recall of the words in any order; examples of free learning are shown by the work of Bousfield (e.g., Bousfield, Cohen, & Whitmarsh, 1958) and Deese (1961). In this area several variables have been isolated and manipulated with college students; namely, interitem associative strength, meaningfulness, intralist similarity, and length of list. Aside from the series of studies emanating from the Catholic University of America which will be considered at a later point, only a handful of studies

involving children remain in this important and active research area.

The influence of interitem associative strength has not been investigated with children. The fact that free-association norms are not readily available for children may account for this neglect; fortunately, a forthcoming publication (Palermo & Jenkins, 1963) will give word associates for many age levels. Considering the importance that some have placed on interitem associative strength (Deese, 1961), an extension of this work to children, for whom interitem associations may be weaker, may produce results which differ from those obtained with college students.

With regard to meaningfulness, an older study by Achilles (1920) using children (aged  $8\frac{1}{2}$ – $11\frac{1}{2}$ ) and a 5-minute study period reported greater recall for words ( $\bar{X}=5.22$ ) than for nonsense syllables ( $\bar{X}=1.86$ ). The basic results are in line with those of Postman and Phillips (1961) who, with college students, obtained greater recall for adjectives than for nonsense syllables in a free-recall situation.

The single study of formal intralist similarity in free recall with children (Horowitz, 1961) was performed with high school students. For the first four alternate study and free-recall trials the high similarity condition was superior to the low similarity condition. The nearest comparison with college students is found in an experiment by Miller (1958) in which the free recall of redundant strings of letters was studied, the more redundant strings, i.e., items with high formal intralist similarity, being more easily learned. The same relationship has been found for meaningful similarity (Underwood, Runquist, & Schulz, 1959) where the free learning of high and low similarity adjectives resulted in better recall of the highly similar items.

The work of Murdock (1960), in

which the length of list in free learning was manipulated, has not been studied with children; however, there is no particular reason to expect the linear relation between the number of items learned and the total presentation time which he found with college students will not hold for children.

A widely studied phenomenon in free learning is that of associative clustering in recall. First studied by Bousfield (1953), associative clustering has only recently been investigated in children. Bousfield, Esterson, and Whitmarsh (1958) compared third-grade, fourth-grade, and college subjects in the type of clustering used (perceptual or conceptual) and in the degree of clustering, by presenting 25 stimulus pictures which could be grouped on the basis of five colors or five categories such as birds, fruits, flowers, nature, and vegetables. A direct relationship was obtained between the degree of total clustering for the color and conceptual categories combined and the age of the subjects. Holroyd and Holroyd (1961), using high school students, compared the degree of clustering and total recall for lists differing in the degree of associative frequency, that is, the frequency with which words were elicited by four key words or clusters. Both degree of clustering and the amount of absolute recall were directly related to the associative frequency of the words represented in the four categories, thus replicating the findings with college subjects by Bousfield, Cohen, and Whitmarsh (1958).

#### PAIRED-ASSOCIATE LEARNING

It has been considered both logically and theoretically convenient to view paired-associate learning as occurring in two general stages, the response-learning stage and the associative or hook-up stage (e.g., Underwood, Runquist, & Schulz, 1959). Spiker (1960b) has also recognized the advantages of the paired-

associate method which allows the possibility of an independent manipulation of stimulus and response functions and is easily adapted for children; however, in general, investigators using children have not taken advantage of the analytical properties that this method provides.

*Backward Associations.* One aspect of the association studied with college subjects is the formation of a backward or R-S association in addition to the usually studied forward or S-R association (e.g., Richardson, 1960). Similarly, backward associations have been demonstrated in children (Grades 4 and 6) by Palermo (1961) who estimated the R-S association to be approximately 57% as strong as the S-R associations; no difference was obtained between R-S associations for the two grades.

*Meaningfulness.* In spite of recent theoretical interest in this variable (e.g., Underwood & Schulz, 1960), there are no studies reported in which meaningfulness was specifically manipulated as a task variable with children as subjects. If the predifferentiation (PD) procedure may be considered as a means by which meaningfulness may be "built-in," several studies using children have been reported which represent the manipulation of stimulus meaningfulness. Saltz, Metzen, and Ernstein (1961) gave two groups of children (aged 6-7 and 10-12) 15 trials of PD training in which the stimuli, four silhouettes of common objects, were associated with stars placed randomly on a board. Following PD training the four PD stimuli and four control stimuli were combined with eight single-letter responses to form the paired-associate task. Comparisons of the PD and control pairs revealed a slight inhibitory effect for both age groups, the effect being greater for the older subjects.

Two experiments by Muehl (1960,

1961) were concerned with the effect of visual discrimination pretraining upon subsequent paired-associate learning. In the first experiment (Muehl, 1960), pretraining consisted of a matching task in which the subject matched a visually presented stimulus with the same item appearing on a response card. Three groups of kindergarten children differed in the type of item that was matched, one group matching the four words which would form the subsequent learning task, one group matching four completely different words, and one group matching a series of four geometrical forms. Following six matching trials the subjects were confronted with four paired associates, the stimuli being the four words of the first group and the responses being the verbal pronunciation of the words. The subjects attempted to anticipate the verbal response within the 2-second interval preceding the presentation of the stimulus and the experimenter's pronunciation of the word. The group receiving the same stimuli for matching and learning gave significantly more correct anticipations for the 12 learning trials than the other two groups, thus indicating a positive effect for stimulus PD. In a later experiment, Muehl (1961) found that PD training with identical or similar stimuli produced no differences in paired-associate learning.

These experiments with children indicate that both positive and negative effects of PD training may be obtained. In this respect the results reported with adults are comparable, stimulus PD training producing both positive (e.g., Gannon & Noble, 1961) and negative (e.g., Underwood & Schulz, 1960) effects. In view of the contradictions in the results for both adults and children, further experiments may be necessary to resolve these discrepancies (in this regard see Schulz & Tucker, 1961).

Research with college subjects has

shown that manipulation of response meaningfulness of paired associates produces reliable positive effects in learning (Underwood & Schulz, 1960). A suggestive comparison experiment with children is found in an older study by Stoddard (1929), using high school students, who showed almost two times as much learning for 25 French-English vocabulary pairs than for the "turned-over" English-French pairs, indicating the large effect of response meaningfulness on paired-associate learning. It is possible that future research will reveal that the various measures of meaningfulness may not predict learning as well for children as for adults. For example, pronunciation ratings of nonsense syllables may lose their predictive value for the learning of nonreading subjects. At any rate, manipulations of this variable would appear to offer possibilities for a fruitful series of comparative adult-children studies.

*Intralist Similarity.* As in free learning, the three types of intralist similarity, formal (letter duplication), meaningful (synonyms), and conceptual, have all been shown to be influential variables in paired-associate learning of adults. In spite of the importance of this variable for the learning of college subjects, no studies have been reported which directly attack intralist similarity with children. An inspection of the pairs employed by Postman and Murphy (1943) reveals high intrastimulus and response similarity being produced by synonyms and conceptual similarity among the experimental pairs and low similarity among the control pairs. These pairs were combined in a single list and learned by eighth-grade students. The significant difference in speed of learning in favor of the control pairs could be taken as evidence of the operation of intralist similarity of children. However, this type of "inferential"

manipulation of similarity is only suggestive and cannot replace experiments in which similarity is expressly manipulated. It would be valuable, for example, to determine if children demonstrate similarity effects for the three types of similarity. It is possible that for the young subject, for whom the meaning of words and concepts is not known, only manipulations of *formal* similarity will be found to be effective; that is, an interaction may be expected to occur between type of similarity manipulation and age of the subject.

*Associative Connection.* Manipulation of the strength of the pre-experimental associative connection has been found to be an effective variable for adult learning. For example, Key (1926) found immediate recall scores to be directly related to the strength of the associative connection as indexed by the Kent-Rosanoff norms. Several experiments have been reported in which variations in the strength of the associative connection have been studied with children. In a study by Akutagawa and Benoit (1959) three lists of paired pictures were constructed, one list of familiar associates, a second of arbitrary associations, and a third of interfering associations; the subjects ranged in age from 8 to 13. The mean number of correct anticipations varied directly with the assumed strength of the S-R association. This study must be viewed with caution, however, since the three lists were presented to all subjects in the same order, easy to difficult, thus possibly confounding lists and practice effects.

In a more adequately designed experiment, Castaneda, Fahel, and Odom (1961) obtained word associations from a population of children (Grades 4-6). Two lists were constructed from the 63 stimulus adjectives, each of the six pairs having the same stimuli but differing in

the strength with which the responses were elicited in the word association test. Two new groups of subjects (Grades 5 and 6) learned either the high or low association list to a criterion of one perfect or for a maximum of 15 trials. The difference between the learning measure for the two groups produced a highly significant difference. McCullers (1961) speculated that control of the age of the subjects possibly would produce differential results, since the children would differ in their linguistic and classroom experiences. Using the norms of Castaneda et al., McCullers selected 40 fourth- and 40 sixth-grade subjects and assigned them to one of eight conditions, two lists of high or low associative strength, and four interpair intervals of 2, 3, 4, or 6 seconds. As with the study of Castaneda et al., the two lists produced highly significant variation in the mean trials to learn, while no significant differences between the two grades nor the interitem intervals were obtained.

*Distributed Practice.* In spite of the large number of experiments in which distributed practice has been manipulated, only one study was found in which children served as subjects. Kirkwood (1926) used a paired-associate procedure in which preschool subjects learned to pair 20 pictures with blocks of various geometrical forms. Massed practice consisted of daily training, distributed practice of training every other day; training was continued for both groups until a criterion of three successive correct trials was reached. Subjects in the distributed condition reached the criterion five practice periods earlier than those in the massed condition and showed higher scores on each trial during learning.

While work reported by Underwood (1961) indicates that distributed practice is not an important variable in the *learning* of paired associates, recent

studies (e.g., Underwood, Keppel, & Schulz, 1962; Underwood & Schulz, 1961) demonstrate that under certain conditions, e.g., an A-B, A-C relationship, distributed practice in learning produces marked facilitation in *retention*. To the extent that a child's learning may be described in terms of an A-B, A-C transfer paradigm, this would certainly appear to be a fruitful research area for the educational psychologist.

#### SERIAL LEARNING

Spiker (1960b) pointed out that "whereas the literature on serial verbal learning in adults is massive, the research with children is less impressive [p. 404]." While approximately 24 studies employing the paired-associate technique were discovered in preparing the present review, less than half as many using serial learning was found. Spiker speculated that this state of affairs may derive from the fact that serial learning is too difficult for the elementary school child and, in addition, that the serial task is a relatively unanalytical tool, each item functioning both as a stimulus and as a response.

Of the serial experiments conducted prior to 1940 using children as subjects, most were systematic studies of the memory span, determining its size as a function of age, modes of presentation, and types of material. Although these studies report a slow increase from early age to ages 10-12 and a slower increase up to age 18, little information is offered by way of accounting for these changes. It would seem that developmental studies, per se, are of limited value to the student of verbal learning other than indicating a lower age limit for potential experimental subjects, unless, however, developmental relationships are found to provide the background for differential theoretical predictions. For example, the age of the subject may be one way in which the strength and complexity of

the associative network may be manipulated in a verbal-learning experiment.

Most investigators agree that children, as do adults, produce a bowed serial position curve (e.g., Andrews, Shapiro, & Cofer, 1954; Barnett, Ellis, & Pryer, 1960; Cassel, 1957), while others have observed changes in the shape of the curve as a function of age. Unfortunately, these latter studies (e.g., Lepley, 1934; Peixotto, 1942) represent procedures and results so complicated that it is impossible to give any substantial generalizations in summary form. In addition to these studies, one report was concerned with the effect of an isolated term in serial learning (Tatuno, 1961); however, in this case the use of subjects of differing ages may have produced a confounding variable since elementary, junior high, and college students were placed in independent experimental groups, thus confounding ability level and the effect of the experimental variable.

Little work has been accomplished in serial learning with children, but this may be a reflection of the biases of the verbal-learning experimenter to work with more analytical learning tasks. Recently, an important line of research has been pursued by Young (1961, 1962) with college subjects, in which an attempt has been made to determine the functional stimulus in serial learning. Some degree of the complexity of this problem may derive from the fact that several choices of stimuli are available to the college student, for example, the preceding stimulus, the preceding stimuli, or the serial position, with some students employing one type of stimulus, some another, and some a combination. Research with children may help to untangle these complex relationships since the child may not have developed as strong stimulus-selection biases as the college student. If this is the case, then

a specification of the stimulus in serial learning may be more easily made with children. In any case, to the extent that the work concerned with determining stimulus-response functions in serial learning is successful, the use of the serial task in verbal learning may again be seen.

#### OTHER LEARNING METHODS

*Syntactical Learning.* Research in this area has taken two forms, namely, the study of the effect of syntax on the learning of nonsense sentences (Epstein, 1961) and the study of the learning and retention of connected discourse (e.g., Slamecka, 1959). The work on connected discourse has its counterpart with children in the many early studies on the Ballard-Williams reminiscence effect; however, these experiments are of little value here since interest was not directed toward the study of the learning of connected discourse. Early work by Foster (1928) demonstrated the effect of repetitions on the mean number of words from a story that a child can recall, while Stroud and Maul (1933) found a direct relationship between the amount of poetry recalled and the age of the subjects. Although this type of learning would appear to have much in common with the learning in the classroom, there have been no systematic experiments with children that have been concerned specifically with syntactical learning. That such experiments may be reasonably conducted with children is indicated by the work of Berko (1958) who has demonstrated that strong and consistent syntactical habits have already been acquired by the preschooler.

*Verbal Concept Learning.* Although not exclusively in the area of verbal learning, mention will be made of one study which represents an attempt to extend the findings of adult concept learning to that of children. The work of Underwood and Richardson (1956)

was repeated by Runquist and Hutt (1961) in an attempt to determine the effect of grade level (9-12) on the acquisition of high dominant concepts. The task was to discover the four concepts which would describe 16 stimuli, represented visually as pictures or verbally as words; with the 16 stimuli there were four instances of each concept. Two sets of pictorial stimuli were used, one set which emphasized the dominant concept, the other which minimized it. The mean total correct concept naming in 15 trials was found to increase as a function of grade for all three sets of stimuli, with an indication of an interaction between the sets and grade, the difference between the types of stimuli being greater for Grade 9 than Grade 12.

*Verbal Discrimination.* Several studies utilizing college subjects have been reported (e.g., McClelland, 1942; Runquist & Freeman, 1960) in which a verbal-discrimination procedure was employed; no studies were found in which children were used.

#### TRANSFER

*Learning-to-Learn.* The mere learning of a series of paired-associate or serial lists by college subjects has been shown to produce increased performance on the learning of subsequent and unrelated lists (e.g., Greenberg & Underwood, 1950; Meyer & Miles, 1953). Gates and Taylor (1925) testing children investigated the effect of training in digit memory span; this training produced a gain of 2.07 digits for the experimental subjects following 78 days of practice in memorizing digit series of various lengths, while the control subjects produced a gain of .73 digit. A study by Drees (1941) involved sixth-grade children in a paired-associate task for which names of trains served as stimuli and two-digit numbers as responses. The experimental group, receiving three months of practice associating other

objects with different two-number combinations, improved in subsequent paired-associate learning; however, this must not be taken as a pure case of learning-to-learn transfer since the possibility existed that specific number combinations were repeated during the training period. Using a different technique, Spiker (1960a) demonstrated that training fourth-grade students in the use of mnemonic devices on one list transferred to the learning of a second list. In spite of the fact that these studies demonstrate in a tangential fashion nonspecific improvement in learning, it is evident that the proper comparison experiments have not been reported so that the degree and extent of learning-to-learn with children has yet to be estimated.

*Warm Up.* As another type of nonspecific transfer, warm up has been demonstrated in learning with adults (e.g., Thune, 1950, 1951), but there are no reported studies with children. In an unpublished study under the direction of the author, Richard Piagari manipulated degree of warm up on a serial task with seventh- and eighth-grade children. Following four practice trials, the subjects received either 0, 7, or 10 trials on serial List A. On the second day these groups received, respectively, 10, 3, and 0 trials on List A, followed immediately by List B which was learned to one perfect recitation. Since all three groups were equated on the degree of learning-to-learn, that is, 10 trials on List A and 4 practice trials, any difference observed in learning List B would be a function of warm up. In terms of learning on List B an inverse relationship was obtained between the number of warm-up trials (List A learning on Day 2) and trials to reach criterion on List B. These results directly parallel those reported by Thune (1950) for adult paired-associate learning.

*Specific Transfer.* Although transfer

designs allow manipulation of various similarity relations between stimuli and responses across lists, it has been found useful to consider key paradigms which define the limits of the similarity relations. For example, Twedt and Underwood (1959) studied four transfer paradigms, A-B, C-D; A-B, C-B; A-B, A-C; and A-B, A-Br, the first letters referring to the stimulus terms, the second letter the response terms, and the relationships between stimuli and responses of the two lists referring to the similarity obtaining. The last paradigm, A-Br (omitting the first-list designation), indicates a second list in which the stimuli and responses of the first list are utilized but are repaired in the second list. A recent study (Jung, 1962) would seem to indicate that the amount and direction of transfer in the C-B paradigm is dependent upon the meaningfulness of the response term; that is, with low levels of meaningfulness positive transfer may be produced while with high levels of meaningfulness negative transfer is produced. Jung interpreted this interaction in terms of the transfer of response learning from the first to the second list, low levels of response meaningfulness producing sufficient positive transfer through the transfer of response learning to overcome the negative effect produced by interference from the backward association (Twedt & Underwood, 1959). Presumably the same interaction will be obtained between the amount of transfer and meaningfulness for the A-Br paradigm, so that a statement of the expected ordering of the four paradigms must consider the meaningfulness of the response term in the C-B and A-Br paradigms. For the Twedt-Underwood study, if the C-D condition is taken as the baseline for zero transfer (i.e., control for nonspecific transfer), increasing negative transfer was obtained for the C-B, A-C, and A-Br conditions, respectively. These

findings are in line with Jung's predictions since paired adjectives, which necessitate minimal amounts of response learning, were used as learning materials. Although several investigators have attempted to isolate various factors to account for these transfer results (e.g., Barnes, 1960; Keppel & Underwood, 1962), an attempt will be made here only to determine if these general relationships have been obtained with children.

Matousek (1939) compared the retroactive inhibition (RI) produced by the C-D, C-B, and A-C conditions with fifth- and sixth-grade subjects. The materials consisted of names of men and dates, such as Columbus—1503, familiar combinations being eliminated. The response terms, although unfamiliar number combinations, should probably be considered relatively meaningful. All groups spent 7 minutes studying the list of pairs, followed either by a 2-minute recall test or a recognition test in which the name was presented with five alternative dates. The second list was presented for 7 minutes, followed by a 2-minute test; finally, recall was taken for the first list. For both recall and recognition, the A-C group showed the greatest RI, followed by the C-B, and finally, the C-D group. The percentage RI as measured by the recall method was 17.7%, 8.7%, and 3.4% for the three groups, respectively. To the extent that RI and transfer are related, these results directly replicate with children, the findings of Twedt and Underwood with adults.

Other investigators have studied one or two of the four paradigms with children. Gladis (1960), using paired one-syllable nouns in a C-B paradigm, found negative transfer for third- and seventh-grade children, but zero transfer for fifth-grade subjects. Mean differences between experimental and control groups in the mean trials to reach criterion were

—2.2, .2, and —1.3 for the third, fifth, and seventh grades, respectively. Spiker (1960a) reported a study in which A-C and C-D pairs were combined in a single mixed-list design. Two groups of fourth-grade subjects received either 6 or 15 trials on the first list, followed by a single transfer list. No difference was obtained between the A-C and C-D pairs over the 12 second-list trials for the low degree of first-list learning, while significant negative transfer was observed for the high degree of first-list learning. Finally, Iwahara and Sugimura (1960) have reported the only study in which the A-Br condition was employed with children (aged 15–19); four degrees of first-list learning of paired two-syllable nouns were employed. Negative transfer was obtained for all but two of the comparisons determined for the several response measures, the two deviant comparisons producing only slight positive transfer.

As yet there have been no studies with children in which all four paradigms were represented. Even though three paradigms were studied by Matousek (1939), a true paired-associate technique was not employed; the newest paradigm, A-Br, has been studied only once with children and there is a question as to whether the estimate from first-list learning of nonspecific transfer is the appropriate control comparison to be made. The simple replication of the Twedt-Underwood study with children of various ages would be of interest, since it is possible that factors that have been isolated for adult learning may not be as important for children learning or may be of differential influence. For example, response learning may be a more important factor for children than for adults since the former have not had as much experience with the language from which the learning materials have been derived. If this is the case, systematic studies with these key paradigms may prove to be

quite valuable for the study of verbal learning.

*Interlist Similarity.* No studies have been reported in which either stimulus or response interlist similarity has been manipulated with children.

*Mediation.* Two studies using children have been reported in which an attempt was made to extend the findings of Bugelski and Scharlock (1952) with adult subjects. Norcross and Spiker (1958) studied positive and negative mediation with a mixed-list design in which each subject served in the three mediation conditions. All subjects received a succession of three lists, List 1 consisting of six pairs of pictures, with the construction of Lists 2 and 3 being dictated by the experimental conditions. That is, the relationships obtained across the three lists allow three mediation paradigms to be represented in List 3. Specifically, if the stimuli and responses of the three lists are expressed by appropriate letters, the construction of List 3 produced two pairs which formed a positive mediation paradigm (A-B, B-C, A-C), two pairs a negative paradigm (A-B, B-D, A-C), and two control pairs (A-B, D-C, A-C). Thirty kindergarten children received the three lists. Since these children had limited reading ability, each list consisted of pairs of common pictures for which the subject anticipated the name of the second picture when presented with the first. In the first experiment, significant positive mediation was obtained, while the control and negative mediation conditions did not differ significantly. In the second experiment only the control and negative conditions were studied, with three pairs representing each condition instead of two. Twenty-four first-grade subjects and a faster rate of presentation were employed. These procedures produced significant negative transfer, allowing the conclusion that both negative and posi-

tive mediation may be demonstrated in relatively young children.

Berkson and Cantor (1960) studied the positive mediation paradigm with normal and retarded elementary school children (aged 9-12) to provide more tests of mediation along the age dimension, ranging from kindergarten-first-grade subjects (Norcross & Spiker, 1958) to college subjects (Bugelski & Scharlock, 1952). A design was employed in which one half of the subjects received the mediation condition (A-B, B-C, A-C), the other half of the subjects received the control condition (X-B, B-C, A-C), where the As were odd numbers, the Xs even numbers, the Bs pictures of common objects, and the Cs various colors. Two comments on design may be made: (a) the stimuli, odd numbers for the A stimuli and even numbers for the X stimuli, may represent relatively similar stimulus sets and thus not provide a truly neutral control condition; (b) the relationship between Lists 1 and 3 for the experimental group forms an A-B, A-C relation while for the control group an A-B, C-D relation, thus allowing simple transfer effects between Lists 1 and 3 (negative) to operate against the expected mediational effect (positive). However, since significant positive transfer was obtained, it is evident that any negative effect produced by the second factor was not sufficient to cancel the positive influence of the mediation relation. In an experiment with college subjects, McGehee and Schulz (1961) studied positive and negative mediation, using mediational links inferred from word association norms rather than specific pair learning. In this case, the facilitating effect of the positive mediation paradigm was only strong enough to overcome the negative transfer effect, the experimental and control groups producing the same amount of transfer.

The fact that mediation can occur with relatively young children may provide a useful source of theoretical predictions. One obvious question that may be asked is whether subjects differing in age differ in the strength of, or in the use of the implied mediational chains. In order to answer this question comparisons between the percentage of subjects demonstrating mediation may be made to determine the number of subjects using mediational chains for various age groups. Another question refers to the theoretical expectation that mediational chains, such as A-B-C, should be expected to take a longer time to occur than a direct chain of A-C. It follows that variations in the anticipation times for List 3 should affect the occurrence of mediation, longer times allowing mediation to occur, shorter times blocking or even inhibiting the runoff of these chains. However, if the adult subject has developed strong mediational processes, then it may be difficult to produce this latency effect. On the other hand, if children and adults differ in the strengths or use of mediational chains, and if the strength determines the speed with which these chains are runoff, then an interaction may be expected in which adults are less affected by variations in anticipation times on List 3 than are children.

*Degree of First-List Learning.* Two studies previously discussed, Spiker (1960a) and Iwahara and Sugimura (1960), varied the degree of first-list learning in transfer paradigms. Spiker found more negative transfer in an A-C paradigm with greater first-list learning, while Iwahara and Sugimura reported a curvilinear relationship between amount of negative transfer and degree of first-list learning for the A-Br paradigm. The results of Iwahara and Sugimura do not agree with those recently obtained by Postman (1962) who found an increas-

ing amount of negative transfer as degree of first-list learning is increased. However, since Iwahara and Sugimura did not employ a C-D control, a direct comparison between the two studies is not warranted. In addition, differences in the number of pairs (4 and 10) may have affected the amount of interference generated by the A-Br paradigms for the two experiments.

*Summary.* A relatively large number of studies have involved transfer paradigms with children as subjects; all four key paradigms and various mediational designs are represented, producing phenomena which have been observed with adult subjects. Investigations of learning-to-learn, warm up, and interlist similarity in children have not been reported. Future work with children should perhaps be directed towards the goal of further isolation of factors thought to be operating in these transfer and mediation designs. Adult subjects may represent learning in which several factors are operating concomitantly, so that the study of children may allow for a more "pure" case of whatever factors are theoretically assumed to be occurring in adults.

#### SIMPLE RETENTION

The phenomenon of simple retention refers to the operations of taking recall at various intervals following the learning of a single task. Research in this area has been limited by various methodological problems. For example, Underwood (1957) demonstrated that the use of the counterbalanced design in the study of retention introduces the problem of proaction from previously acquired lists interfering at recall. Similarly, problems of degree of learning (Underwood, 1954b) and of intra- and interlist similarity relations (Underwood, 1954a) have posed experimental and measurement problems for researchers in this area. Experimental work with

children has not been affected by these problems for the simple reason that, in general, independent groups have been used and few variables have been manipulated. For the subject naive to verbal-learning experiments, any forgetting observed has been thought to be due largely to interference from well-learned language habits. For example, recent statements of interference theory (e.g., Postman, 1961; Underwood & Postman, 1960) have leaned heavily on the premise that much of verbal learning involves the unlearning of previously learned, stronger letter- or word-language habits. Viewed in this way, the difficulty in learning a low meaningful list of nonsense syllables may be thought of as requiring the temporary modification of the verbal repertoire the subject brings with him to the laboratory, and the forgetting observed over 24 hours being due, in part, to the regeneration or spontaneous recovery of these stronger language habits which interfere with recall. If it is assumed that the child has a less complicated and less practiced associative network, then less forgetting may be expected for children than for college students. Unfortunately, no available data allow a test of this expectation. The evidence presented below represents, with one exception, retention values for children roughly approximating the 80%–90% obtained from college subjects (e.g., Underwood & Keppel, 1962). The retention of children in these studies might in fact be expected to be less than the values obtained for college subjects. The reason for this is that the low degrees of learning employed in the children studies may result in associations which are more susceptible to interference than those produced by the higher levels of learning employed in the studies with college subjects.

The data on the retention of children

noted above come from a series of experiments from the Catholic University of America which allows the determination of the retention of children naive to verbal-learning experiments. Lahey (1937) reported for control groups studying a list of 25 verbs for 5 minutes, 24-hour retention values ranging from 72%–82%. Houlahan (1937), using the same materials as Lahey, reported control-group retention of 96% after 17 minutes and 85% after 24 hours. McDade (1943) had the subjects study a list of 20 objects for 30 seconds, with retention for the names of the objects being taken at various intervals. For younger subjects ( $MA=9.5$ ), 93% retention was observed and for older subjects ( $MA=12.5$ ), 89% retention was obtained after 24 hours. Finally, Sharpe (1952) studied the retention of the meaning of a prose passage by administering a series of questions to various groups; 87% retention was obtained after 24 hours. One contradiction to these data is a study by Otto (1961) in which five geometrical form-nonsense syllable pairs were learned by children from several grade levels (Grades 2, 4, and 6). Retention measures taken 24 hours after the attainment of one perfect recitation revealed approximately 63% to be retained. Clearly, in view of the wide differences in the degree and methods of learning and in the materials used represented in the above experiments, systematic studies in which degree of learning is held constant across age levels will be necessary to determine if the expectation of less forgetting by children is to be adequately tested.

Two studies with children have involved the manipulation of a subject variable in retention. Gregory and Bunch (1959) compared slow and fast subjects selected from a population of elementary school children. Fast subjects were drawn from the upper quartile

of the distribution, slow subjects from the lower quartile. Each group of 40 subjects was subdivided into two smaller groups, one being tested immediately after learning, the other 24 hours after learning. The pairs consisted of 10 geometrical drawings paired with the numbers 1-10, a response discovery technique being used; learning was taken to the criterion of one perfect recitation. The fast learners tended to recall more numbers after 24 hours than the slow learners, but the difference was not significant. In addition, since degree of learning for the two ability groups is not equated in this comparison, it is not clear to what extent ability level produced differences in retention (see Underwood, 1954b). Specifically, the mean errors after 24 hours for the fast and slow groups were 2.8 and 4.1, respectively, while the two immediate retention groups produced mean error scores of .9 and 2.0, respectively. Loss scores, obtained by subtracting the respective 0- and 24-hour groups, produces losses of 1.9 and 2.1 for fast and slow learners, indicating the absence of an effect of ability level on retention. This conclusion is supported by the experiment by Otto (1961) who found no differences in retention as a function of the reading ability of children.

*Summary.* Systematic studies of retention using children as subjects have not been conducted. For the evidence that is available, it would appear that Underwood's (1957) estimate of 10%-20% forgetting over 24 hours is a reasonable one for nonpracticed children, a finding which should be quite heartening to the teacher who, in reading introductory psychology tests, has been told to expect 60%-70% forgetting overnight. Aside from the practical implications that research on retention has for education, the study of retention in children would appear to offer the verbal-learn-

ing theorist potential tests of current interference theory.

#### RETROACTIVE AND PROACTIVE INHIBITION

Although retroactive inhibition (RI) has received a great deal of attention since 1900, little systematic research has been done with children. For example, a recent review by Slamecka and Ceraso (1960) of RI and PI studies since 1940, does not mention specifically any studies in which children were used as subjects. Whitely (1927) was the first to investigate RI in children, finding more RI with high school and grammar school subjects than with college students; however, since the experimental conditions were not the same for all groups, these comparisons must be viewed with caution. Several investigators have reported an inverse relation between degree of RI and age and IQ (e.g., Houlihan, 1937; Lahey, 1937; Matousek, 1939). Since degree of RI depends upon both degrees of original learning (OL) and interpolated learning (IL), and since learning rate appears to be a direct function of age or IQ, these studies provide little by way of explaining the obtained inverse relationship between age and IQ and RI. An experiment in which degree of learning is held constant for all age groups has not been accomplished. Therefore, the relationship between RI and age cannot be interpreted to mean that younger subjects are more susceptible to interference until the degree of learning problem has been solved.

The studies that have manipulated task variables in an RI design have produced results which corroborate those obtained with adult subjects. Matousek (1939) demonstrated that the percentage RI depends upon the similarity relationships between the two experimental lists. Lahey (1937) found percentage RI to increase as the IL study time in-

creased from 4 to 14 minutes. Houlahan (1937) varied the temporal position of the interpolated list within a constant OL-recall interval of 17 minutes, finding maximum RI when the interpolated list was presented immediately before recall of the first list; this effect replicated the findings with college subjects by McGeoch (1933a, 1933b) who found maximum RI when IL occurred immediately before recall of the first list. The results of Deese and Marder (1957) who, with high school students, found the number of between-list errors to increase from a 4-minute interval to a 48-hour interval is consistent with the notion of differentiation (Underwood, 1945) in which the subject's discrimination between OL and IL is predicted to decrease as the IL-RL interval is increased. Finally, Cassel (1957) found no difference in the amounts of RI between aments and children matched in MA.

#### FINAL COMMENTS

A review of the available literature on children's verbal learning has revealed a surprising correspondence in the effective variables reported and the relationships that have been identified with adult subjects. Although suggestions for further study have been indicated at various points in this review, a number of comments with regard to the future use of children in verbal-learning experiments seems appropriate at this time.

First, the developmental study, per se, is probably of little value to the verbal-learning theorist unless differential results are to be expected by theory. This is not to say that all research involving children's verbal learning must be directly tied to theory, but that the mere replication of adult studies may offer little to theory directly. To know that the memory span increases with chronological age is certainly of interest to the educational and developmental psycholo-

gist, but unless this relationship can be translated into extant theories and produce predictions that will provide reasonable tests of these theories, the finding is of no immediate use to the verbal-learning theorist. Deese (1952) has said this another way when referring to research of the relative learning abilities of children,

certainly there is very little novel or new information which comes out of this research. Perhaps the next few years may see a revival in interest in this field, with an emphasis upon the *causes* of individual differences with age rather than the simple measurement of such individual differences [p. 297].

Second, research with children may provide information with regard to *new* or *raw* learning, that is, learning which may be considered to have occurred for the first time, rather than being based on previously learned mediators or associations. For example, many adult subjects, in describing the learning of a paired-associate list, report the use of mediational links between the stimulus and the response, rather than a direct association between them (e.g., Bugelski, 1962). This would seem to indicate that the learning observed in these experiments involves largely a transfer of past learning rather than the acquisition of a *de novo* association. But these mediators must have been acquired at some time in the history of the subject and research with children may shed some light on this problem.

Finally, the extension of theories based on the learning of college students to the learning of children would seem both an important and a necessary step in the field of verbal learning. This point is not a contradiction of the first point, however, since the issue here involves tests of theory and application of findings based on college students to younger and possibly older members of the same species.

## REFERENCES

- ACHILLES, EDITH M. Experimental studies in recall and recognition. *Arch. Psychol.*, N. Y., 1920, 6(No. 44).
- AKUTAGAWA, D., & BENOIT, E. P. The effect of age and relative brightness on associative learning in children. *Child Develpm.*, 1959, 30, 229-238.
- ANDREWS, T. G., SHAPIRO, S., & COFER, C. N. Transfer and generalization of the inhibitory potential developed in rote serial learning. *Amer. J. Psychol.*, 1954, 67, 453-463.
- BARNES, JEAN M. "Fate" revisited. Unpublished doctoral dissertation, Northwestern University, 1960.
- BARNETT, C. D., ELLIS, N. R., & PRYER, MARGARET W. Serial position effects in superior and retarded subjects. *Psychol. Rep.*, 1960, 7, 111-113.
- BERKO, JEAN. The child's learning of English morphology. *Word*, 1958, 14, 150-177.
- BERKSON, G., & CANTOR, G. N. A study of mediation in mentally retarded and normal school children. *J. educ. Psychol.*, 1960, 51, 82-86.
- BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, 49, 229-240.
- BOUSFIELD, W. A., COHEN, B. H., & WHITMARSH, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychol. Rep.*, 1958, 4, 39-44.
- BOUSFIELD, W. A., ESTERSON, S., & WHITMARSH, G. A. A study of developmental changes in conceptual and perceptual associative clustering. *J. genet. Psychol.*, 1958, 92, 95-102.
- BROWN, R., & BERKO, JEAN. Word association and the acquisition of grammar. *Child Develpm.*, 1960, 31, 1-14.
- BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, 63, 409-412.
- BUGELSKI, B. R., & SCHARLOCK, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, 44, 334-338.
- CASSEL, R. L. Serial verbal learning and retroactive inhibition in aments and children. *J. clin. Psychol.*, 1957, 13, 369-372.
- CASTANEDA, A., FAHEL, LEILA S., & ODOM, R. Associative characteristics of sixty-three adjectives and their relation to verbal paired-associate learning in children. *Child Develpm.*, 1961, 32, 297-304.
- DEESE, J. *The psychology of learning*. New York: McGraw-Hill, 1952.
- DEESE, J. From the isolated verbal unit to connected discourse. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 11-31.
- DEESE, J., & MARDER, V. J. The pattern of errors in delayed recall of serial learning after interpolation. *Amer. J. Psychol.*, 1957, 70, 594-599.
- DREES, V. The effects of practice on memory performance. *Cath. U. Amer. educ. Res. Monogr.*, 1941, 12(No. 4).
- EPSTEIN, W. The influence of syntactical structure on learning. *Amer. J. Psychol.*, 1961, 74, 80-85.
- ERVIN, SUSAN M. Changes with age in the verbal determinants of word-association. *Amer. J. Psychol.*, 1961, 74, 361-372.
- FOSTER, JOSEPHINE C. Verbal memory in the preschool child. *J. genet. Psychol.*, 1928, 35, 26-44.
- GANNON, D. R., & NOBLE, C. E. Familiarization (*n*) as a stimulus factor in paired-associate verbal learning. *J. exp. Psychol.*, 1961, 62, 14-23.
- GATES, A. I., & TAYLOR, GRACE A. An experimental study of the nature of improvement resulting from practice in a mental function. *J. educ. Psychol.*, 1925, 16, 583-592.
- GLADIS, M. Grade differences in transfer as a function of the time interval between learning tasks. *J. educ. Psychol.*, 1960, 51, 191-194.
- GREENBERG, RUTH, & UNDERWOOD, B. J. Retention as a function of stage of practice. *J. exp. Psychol.*, 1950, 40, 452-457.
- GREGORY, S. C., JR., & BUNCH, M. E. The relative abilities of fast and slow learners. *J. gen. Psychol.*, 1959, 60, 173-181.
- HOLROYD, R. C., & HOLROYD, JEAN C. Associative clustering in a retroaction paradigm. *J. gen. Psychol.*, 1961, 64, 101-104.
- HOROWITZ, L. M. Free recall and ordering of trigrams. *J. exp. Psychol.*, 1961, 62, 51-57.
- HOULAHAN, F. J. Retroactive inhibition as affected by the temporal position of interpolated learning activities in elementary school children. *Cath. U. Amer. educ. Res. Monogr.*, 1937, 10(No. 3).
- IWAHARA, S., & SUGIMURA, T. Associative interference as a function of degree of prior training in paired-associate learning. *Psychol. Rep.*, 1960, 6, 243-246.
- JUNG, J. R. The effects of response meaningfulness (*M*) on transfer of training under two different paradigms. Unpublished doctoral dissertation, Northwestern University, 1962.
- KEPPEL, G., & UNDERWOOD, B. J. Retroactive inhibition of R-S associations. *J. exp. Psychol.*, 1962, 64, 400-404.

- KEY, CORA B. Recall as a function of perceived relations. *Arch. Psychol.*, N. Y., 1926, 13(No. 83).
- KIRKWOOD, JULIA A. The learning process in young children: An experimental study in association. *U. Ia. Stud. child Welf.*, 1926, 3(No. 6).
- LAHEY, M. F. L. Retroactive inhibition as a function of age, intelligence, and the duration of the interpolated activity. *Cath. U. Amer. educ. Res. Monogr.*, 1937 10(No. 2).
- LEPLEY, W. M. Serial reactions considered as conditional reactions. *Psychol. Monogr.*, 1934, 46(1, Whole No. 205).
- MCCLELLAND, D. C. Studies in serial verbal discrimination learning: I. Reminiscence with two speeds of pair presentation. *J. exp. Psychol.*, 1942, 31, 44-56.
- MCCULLERS, J. C. Effects of associative strength, grade level, and interpair interval in verbal paired-associate learning. *Child Develpm.*, 1961, 32, 773-778.
- MCDADÉ, MARY T. F. The relation of the time-interval and of intelligence to reminiscence. *Cath. U. Amer. educ. Res. Monogr.*, 1943, 13(No. 4).
- MCGEEHEE, NAN E., & SCHULZ, R. W. Mediation in paired-associate learning. *J. exp. Psychol.*, 1961, 62, 565-570.
- MCGEOCH, J. A. Studies in retroactive inhibition: I. The temporal course of the inhibitory effects of interpolated learning. *J. gen. Psychol.*, 1933, 9, 24-43. (a)
- MCGEOCH, J. A. Studies in retroactive inhibition: II. Relationships between temporal point of interpolation, length of interval, and amount of retroactive inhibition. *J. gen. Psychol.*, 1933, 9, 44-57. (b)
- MCGEOCH, J. A. *The psychology of human learning*. New York: Longmans, Green, 1942.
- MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
- MATOUSEK, MARY A. Reproductive and retroactive inhibition as a function of similarity in the recall and recognition of paired associates. *Cath. U. Amer. educ. Res. Monogr.*, 1939, 12(No. 1).
- MEYER, D. R., & MILES, R. C. Intralist-interlist relations in verbal learning. *J. exp. Psychol.*, 1953, 45, 109-115.
- MILLER, G. A. Free recall of redundant strings of letters. *J. exp. Psychol.*, 1958, 56, 485-491.
- MUEHL, S. The effects of visual discrimination pretraining on learning to read a vocabulary list in kindergarten children. *J. educ. Psychol.*, 1960, 51, 217-221.
- MUEHL, S. The effects of visual discrimination pretraining with word and letter stimuli on learning to read a word list in kindergarten children. *J. educ. Psychol.*, 1961, 52, 215-221.
- MUNN, N. L. Learning in children. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954. Pp. 374-458.
- MURDOCK, B. B., JR. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, 60, 222-234.
- NORCROSS, KATHRYN J., & SPIKER, C. C. Effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, 55, 129-134.
- OTTO, W. The acquisition and retention of paired-associates by good, average, and poor readers. *J. educ. Psychol.*, 1961, 52, 241-248.
- PALERMO, D. S. Backward associations in the paired-associate learning of fourth and sixth grade children. *Psychol. Rep.*, 1961, 9, 227-233.
- PALERMO, D. S., & JENKINS, J. J. *Word association norms: Grade school through college*. Minneapolis: Univer. Minnesota Press, 1963.
- PEIXOTTO, HELEN E. Intraserial inhibition as measured by reproduction. *J. exp. Psychol.*, 1942, 31, 17-34.
- PETERSON, J. Learning in children. In C. Murchison (Ed.), *Handbook of child psychology*. Worcester: Clark Univer. Press, 1933. Pp. 417-481.
- POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 152-179.
- POSTMAN, L. Transfer of training as a function of experimental paradigm and degree of first-list learning. *J. verbal Learn. verbal Behav.*, 1962, 1, 109-118.
- POSTMAN, L., & MURPHY, G. The factor of attitude in associative memory. *J. exp. Psychol.*, 1943, 33, 228-238.
- POSTMAN, L., & PHILLIPS, LAURA W. Studies in incidental learning: IX. A comparison of the methods of successive and single recalls. *J. exp. Psychol.*, 1961, 61, 236-241.
- RICHARDSON, J. Comparison of S-R and R-S learning of paired-associates. *Psychol. Rep.*, 1960, 7, 225-228.
- RUNQUIST, W. N., & FREEMAN, MADELYN. Roles of association value and syllable familiarization in verbal discrimination learning. *J. exp. Psychol.*, 1960, 59, 396-401.

- RUNQUIST, W. N., & HUTT, VALORIE H. Verbal concept learning in high school students with pictorial and verbal representation of stimuli. *J. educ. Psychol.*, 1961, 52, 108-111.
- RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr66216, Office of Naval Research.
- SALTZ, E., METZEN, J. D., & ERNSTEIN, E. Predifferentiation of verbal stimuli in children. *Psychol. Rep.*, 1961, 9, 127-132.
- SCHULZ, R. W., & TUCKER, I. F. Stimulus familiarization in paired-associate learning. Paper read at Psychonomic Society, New York, September 1961.
- SHARPE, J. F. The retention of meaningful material. *Cath. U. Amer. educ. Res. Monogr.*, 1952, 16(No. 18).
- SLAMECKA, N. J. Studies of retention of connected discourse. *Amer. J. Psychol.*, 1959, 72, 409-416.
- SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, 57, 449-475.
- SPIKER, C. C. Associative transfer in verbal paired-associate learning. *Child Develpm.*, 1960, 31, 73-87. (a)
- SPIKER, C. C. Research methods in children's learning. In P. H. Mussen (Ed.), *Handbook of research methods in child development*. New York: Wiley, 1960. Pp. 374-420. (b)
- STODDARD, G. D. An experiment in verbal learning. *J. educ. Psychol.*, 1929, 20, 452-457.
- STROUD, J. B., & MAUL, RUTH. The influence of age upon learning and retention of poetry and nonsense syllables. *J. genet. Psychol.*, 1933, 42, 242-250.
- TATUNO, T. On the effect of the isolated term and the identical term in serial learning. *Jap. psychol. Res.*, 1961, 3, 17-27.
- THUNE, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, 40, 423-438.
- THUNE, L. E. Warm-up effect as a function of level of practice in verbal learning. *J. exp. Psychol.*, 1951, 42, 250-256.
- TWEDT, HELEN M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, 58, 111-116.
- UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, 59(3, Whole No. 273).
- UNDERWOOD, B. J. Intralist similarity in verbal learning and retention. *Psychol. Rev.*, 1954, 61, 160-166. (a)
- UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, 51, 276-282. (b)
- UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, 64, 49-60.
- UNDERWOOD, B. J. Ten years of massed practice on distributed practice. *Psychol. Rev.*, 1961, 68, 229-247.
- UNDERWOOD, B. J., & KEPPEL, G. An evaluation of two problems of method in the study of retention. *Amer. J. Psychol.*, 1962, 75, 1-17.
- UNDERWOOD, B. J., KEPPEL, G., & SCHULZ, R. W. Studies of distributed practice: XXII. Some conditions which enhance retention. *J. exp. Psychol.*, 1962, 64, 355-363.
- UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, 67, 73-95.
- UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, 51, 229-238.
- UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, 58, 70-78.
- UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Chicago, Ill.: Lippincott, 1960.
- UNDERWOOD, B. J., & SCHULZ, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. *J. exp. Psychol.*, 1961, 61, 228-235.
- WHITELY, P. L. The dependence of learning and recall upon prior intellectual activities. *J. exp. Psychol.*, 1927, 10, 489-508.
- WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.
- YOUNG, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, 74, 517-528.
- YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.*, 1962, 63, 307-313.

(Received August 13, 1962)

# Psychological Bulletin

## SUBLIMINAL STIMULATION:

### A PERVERSIVE PROBLEM FOR PSYCHOLOGY<sup>1</sup>

WILLIAM BEVAN

*Kansas State University*

After a brief historical introduction and a discussion of methodological issues, this substantive review concentrates upon the relevant literature of the past 5 years. Special attention is given to the writer's own experiments on the effect of subliminal anchors upon psychophysical judgments. (86 ref.)

A review of the history of modern psychology yields relatively few general principles of marked significance. Among these few there are two the importance of which is generally recognized: mental events are capable of measurement and behavior may be determined by immediate variables of which the behaving individual is essentially unaware. The first of these is identified with Fechner. Though the notion of measurement of mental phenomena goes back at least as far as Herbart and Weber, it was Fechner who first formulated a general expression for the psychophysical relationship and who codified a set of procedures by which this could be tested. The second, the notion of the unconscious, is identified with Freudian psychology. Unlike the former, which reflects the rational approach of nineteenth century laboratory science, it resulted from clinical intuition and has failed general acceptance among psychologists until the relatively recent past. Indeed, theoretically oriented experimental in-

vestigations of the unconscious date only from the past 1 or 2 decades.

Meanwhile, curiosity about the possible influence of subthreshold stimulation has been in evidence for some time. According to Baker (1937), the earliest experiment is that of Suslowa, who studied the effect of weak electrical stimulation upon the two-point threshold. Unfortunately, a variety of situations has been used in these studies and the definition of subliminal stimulation remains unclear. In some, the critical stimuli are ostensibly below the limits of detectability (absolute threshold). An example of this sort of experiment is the recent study of Black and Bevan (1960) in which subliminal stimuli were introduced prior to the presentation of supraliminal series stimuli in order to determine possible effects of the former upon the latter. In others, the critical stimuli closely resemble or are masked by other elements of the stimulus complex such that confident discrimination is not possible. This type of study may be illustrated by Day's (1956) study of serial nonrandomness in the auditory differential threshold. Finally, in a third class of situations, the critical stimuli are available for perception but lie outside the field of attention. Here the

<sup>1</sup> This review was prepared under Contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research. It is part of a project entitled, "The Evolution of Perceptual Frames of Reference."

verbal conditioning study of Greenspoon (1955) provides a convenient example.

*The "New Look" Approach to Subliminal Perception.* Interest in subliminal perception during the last decade and a half has been dominated by a concern for perceptual defense. The pioneer experiment on this phenomenon is the McGinnies (1949) "dirty-word" experiment. In this study taboo words had elevated recognition thresholds and evoked GSR responses prior to the correct report of them by the subject. A variety of criticisms has been made of this and similar studies. It has been argued that the difference in threshold between neutral and taboo words reflects a difference in familiarity (Solomon & Howes, 1951) and that, in the case of taboo words, the subject may be *suppressing* rather than *repressing* his response. An attempt to untangle these problems has given rise to experiments like that of Lazarus and McCleary (1951) who simulated the anxiety response associated with the taboo words by a pretest application of conditioning shock along with the presentation of certain nonsense syllables. The subsequent recognition thresholds for these taboo syllables were reliably higher than for other neutral syllables, although there was no difference in familiarity among the syllables nor was there any reason to expect suppression.

The theoretical antithesis of perceptual defense is selective sensitization. If anxiety is expected to raise thresholds, positive values are said to lower them. Thus, McClelland and Liberman (1949) found that subjects with high need achievement scores on the TAT had low recognition thresholds for achievement-related words, while low achievers more quickly recognized security-related words. But the results on sensitization, like those on defense, are not completely consistent. For example, Gilchrist, Lude-

man, and Lysak (1954) reported an increase in threshold for both positively and negatively toned words; and, indeed, there is much everyday evidence to support the common-sense expectation that, under certain circumstances, the appropriate response to a threatening stimulus is vigilance and a lowered threshold. A major problem lies in differentiating the circumstances calling for defense, on the one hand, and vigilance on the other. It is interesting to speculate about the difference between stimuli that threaten the psychological integrity of the perceiver—his own good impression of himself and his physical integrity—his safety or comfort. This is, perhaps, a difference between anxiety and fear or between panic and pain. One may anticipate that fear or pain, in contrast to anxiety or panic, are associated with vigilance. Certain of the data to be described later appear to be in line with this supposition.

Eriksen (1952) early emphasized the importance of considering between-subject differences in the design of experiments on subliminal stimulation. An experiment by Carpenter, Weiner, and Carpenter (1956) illustrated this importance. Subjects classified as sex sensitizers, or as hostile, perceived sex-related or hostile stimulus words more readily than sex repressors or nonhostile subjects. Recognition of the importance of individual differences points to a source of weakness in many studies of subliminal stimulation. Common practice has been to identify as subliminal any stimulus magnitude that is less than either the absolute or the differential threshold. But a precaution is in order. The threshold is a statistical estimate and for a stimulus to be consistently subliminal, it must lie outside of the population of values from which the threshold is derived. A single supraliminal response to a presumed subliminal stimu-

lus forces recognition of the fact that some indeterminate number of failures to perceive it are false negative responses. And, as Goldiamond (1958) and Eriksen (1956b) observed, if false failures at recognition occur in sufficient number, they will create a false impression of a subliminal effect.

#### METHODOLOGICAL CONSIDERATIONS

From its inception, the recent work on perceptual defense has been plagued by methodological problems. It was early objected that threshold differences between taboo and neutral words reflected differences in familiarity rather than defense (e.g., Howes & Solomon, 1950). More fundamental is the question of perceptual defense versus response suppression. For some time, the best that has been done has been to subject the subject to posttest interrogation about his experience. Meanwhile, Garner, Hake, and Eriksen (1956) have proposed the principle of converging operations as a device for differentiating among intervening processes and have applied this to the problem of subception. To demonstrate a perceptual discrimination below the level of awareness, several experiments are necessary, even if one equates consciousness with verbal report. Under these circumstances it is necessary to demonstrate a higher correlation between stimulation and GSR (or some other physiological index) than between stimulation and all verbal response systems that can be identified.

Edwards (1960) has shown that recognition thresholds for tachistoscopically presented words vary with the response method employed (multiple-choice, discerning guess, and free verbalization) and has concluded that subliminal perception is a methodological artifact. Granting a relationship between form of response and recognition threshold, Edwards' conclusion is not

warranted unless it is demonstrated that nonverbal response systems fail to correlate with stimulation. Meanwhile, Mathews and Wertheimer (1958) had already proposed a pure measure of defense independent of response suppression. In this, subjects are given a list consisting of an equal number of neutral and emotion-provoking words. These are all identified as words to be tachistoscopically presented. In fact, only half the words listed are presented. By subtracting a performance measure for the "stimulus absent" words (which represent response bias) from the "stimulus present" words (which relate to both perceptual and response tendencies), one arrives at an estimate of "pure" perceptual tendency. A comparison of these difference scores for the neutral and the emotional words is then assumed to indicate the operation of perceptual defense. Mathews and Wertheimer reported such a demonstration for subjects scoring high on the Hysteria scale of the MMPI but failed to find similar evidence in the responses of subjects scoring high on the Psychasthenia scale.

The stereoscope has been used in two experiments (Davis, 1959; Van de Castle, 1960) as a device for studying perceptual defense. The technique involves presenting different material (words or pictures) to each eye and recording what the subject reports. This would seem at first hand to be an excellent solution to the defense-suppression problem, since it deals with supraliminal material. However, it always allows for a neutral alternative in response, and while the perceptually hypersensitive subject may be readily identified, it is not possible to identify the defender as defender or suppressor. Furthermore, the responses are not always "either-or." While the subject may report either the emotionally toned or the neutral stimulus, he may also see both, or parts of

both, or even some integration of the two. Van de Castle (1960) has proposed to resolve this problem by arbitrarily ordering and assigning numerical weights to each of the possible responses. Finally, the subject's response is influenced by eye dominance and it is not possible to correct for this factor simply by counterbalancing the position of the emotionally toned stimulus.

Examination of the literature on subception and perceptual defense reveals no major methodological breakthrough, with the exception of the principle of convergent operations, and this latter has yet to be translated into a particular set of empirical procedures in the service of this area of inquiry. Eriksen (1958), in his Nebraska symposium paper, has performed an exhaustive methodological analysis of the broad spectrum of studies of discrimination without awareness and concluded that there is yet no empirical basis for inferring the existence of a superdiscriminatory system. Rather, he proposes that when the stimulus is extremely weak, the subject may respond to fragmentary cues, with these cues indicating the presence of a stimulus for one response system and not for another. In the case of subception, he suggested that the typical experimental conditions do not allow the subject to state verbally all the discrimination he is capable of making. In a 1962 paper Davis and Hess proposed to measure awareness by rating the subject's responses to open-ended questions following a series of anagram problems. Here, of course, awareness is equated with verbal report and the procedure excludes the possibility of relevant processes that do not fit this category.

#### RECENT LITERATURE ON SUBLIMINAL PERCEPTION

Despite the nagging of methodological issues, a lively interest in subliminal

perception persists to the present. In the past 5 years there have been, in addition to Eriksen's (1958) methodological review, another primarily methodological review (Goldiamond, 1958), several substantive reviews (Adams, 1957; Jenkins, 1957; McConnell, Cutler, & McNeil, 1958), the report of a symposium (Eriksen, 1962), and a monograph (W. P. Brown, 1961). The topic has received special attention in introductory textbooks and in recent texts on motivation (e.g., J. S. Brown, 1961; Hall, 1961). The Brown monograph (1961) comprises a comprehensive treatment of the phenomena associated with perceptual defense, including an extensive substantive review of the literature, a discussion of both the definitional and methodological problems associated with the phenomenon, a critical evaluation of five theoretical approaches, a detailed presentation of one of the author's own experiments, and an exposition of his own theoretical orientation. In the period between Brown's latest literature citation and the summer of 1962 at least 34 experimental studies have appeared.

That the topic has come of age is indicated by several facts. Experiments are no longer simply directed toward establishing subliminal perception as an empirical phenomenon but incline more toward examining its influence upon a variety of behavioral events. In addition, there has been interest in the commercial exploitation of the principle of subliminal perception. In this section we shall be concerned primarily with summarizing the results of this recent empirical literature.

Early in the 1950s interest in subliminal perception was concentrated mainly on perceptual defense. Experimenters treated subliminal perception, for the most part, as the dependent variable and concern centered about the identification of the variables, primarily

stimulus variables, that influenced it. Subliminal perception was identified indirectly by the occurrence of physiological change suggestive of emotional reaction and accompanying a raised recognition threshold for emotion-provoking stimuli. Both the physical and the behavioral properties of stimuli were manipulated as independent variables. Early in the 1960s, interest in subliminal perception appears to have shifted toward its relationship to central—particularly personality—variables and to its role as an independent variable affecting other behavioral processes.

*Stimulus Determinants of Perceptual Thresholds.* Three characteristics of stimuli—their capacity for affective arousal, their familiarity or frequency of occurrence, and their physical intensity—still provoke experimental interest.

Spence (1957) has proposed that anxiety related to threatening stimuli disrupts prerecognition hypotheses near threshold to produce either an elevation or a lowering of the recognition threshold, and his experimental results have indicated no systematic trend toward a defensive, in contrast to a vigilant, reaction to anxiety-related words.

In order to meet methodological criticisms of the Lazarus and McCleary (1951) subception experiment, Dixon (1958b) not only measured recognition thresholds for emotional and neutral words but required his subjects to guess the identity of subliminally presented words, to give free-association responses to these stimuli, and then to match their guesses to the stimuli they thought had been presented. While no reliable increase in latency was noted for the emotional words, the emotional words provoked larger GSRs and five of the seven subjects matched their guesses to the stimuli that accompanied them more frequently than could be expected by chance. Questioning at several stages of

the experiment revealed no evidence of awareness of the subliminally presented stimuli. Fuhrer and Eriksen (1960) repeated Dixon's experiment, but were unable to obtain better than chance matches between presented words and guessed responses. When the experiment was extended to include a group receiving the stimuli at a higher illumination level and another given the stimuli backwards and upside down, better than chance matches were obtained. These latter results, however, would appear to be attributable to structural differences among the stimuli.

Meanwhile, Eriksen, Azuma, and Hicks (1959) reported that subjects can identify the affective quality of a stimulus word reliably when they fail at actual recognition, although there was no evidence of a lower discrimination threshold for affective judgments than for specific identifications. This they took as evidence supporting Eriksen's (1958) concurrent response theory.

Lowenfeld (1961) has objected that the experimental procedure in studies of subception and defense have not met the conditions of accompanying theory, since they have dealt primarily with responses to noxious stimuli or stimuli rendered noxious. More representative studies, he argued, would involve negative affect resulting from the associative connotations of stimuli rather than any noxious quality per se. Accordingly, prior to performing the subception experiment, he induced negative affect into his emotional stimuli by pairing them in a nonsense-syllable exercise with negatively reinforced response syllables. His data supported the occurrence of subliminal perception under these conditions.

Using the technique of stereoscopic perception rather than the threshold technique, Davis (1959) reported that emotionally toned words as well as less

familiar (i.e., less frequently used) words were perceived less frequently when presented simultaneously with neutral or familiar words. There was, however, no significant interaction between these two stimulus variables.

In the Dixon (1958b) experiment described above the methodological goal was to remove response bias associated with the design of the subception experiment. Taylor, Rosenfeldt, and Schulz (1961) attacked the same problem in the design of word frequency experiments with a variant of the forced-choice technique. They required that subjects indicate the spatial position of key words and found, with response bias controlled, frequency of usage to be related to ease of recognition.

In another experiment Dixon (1958a) used the method of closed-loop control to differentiate the perceptual threshold from the threshold for verbal report. In this procedure, the subject kept a continuous threshold setting of a spot of light projected into the left eye. Verbal stimuli, emotional and neutral, were always presented to the right eye at intensities below this value. Generally higher thresholds for the left eye were found to accompany taboo words.

These recent experiments on stimulus determinants of word recognition continue the attempts of earlier studies to clarify the status of subliminal perception as a possible methodological artifact. Although results are sometimes inconsistent, the appearance of the effect appears to persist through each methodological refinement.

*Subliminal Perception as the Perception of Partial Cues.* The view of perception that has emerged from the literature on subliminal processes is that it is an extremely complex response organized in hierarchical fashion, some facets capable of appearance prior to recognition or verbal identification. Mean-

while, as indicated earlier, Eriksen (1956b) has contended that what is identified as subliminal perception in the laboratory may result from restrictions placed by the experimental design on the available number of verbal responses to word stimuli. Boardman (1957) extended this prerecognition hypothesis by arguing that subjects in the subception experiments are capable of discriminating structural differences between tachistoscopically presented shocked and nonshocked words below verbal recognition levels. He further contended that observed differences in GSR reflect this discrimination of elements of partial cues. He demonstrated that almost half of the prerecognition hypotheses concerning taboo words contained two or more letters of the four-letter word presented, while neutral dissimilar words were seldom identified as taboo or taboo-similar words. In contrast, neutral words structurally resembling taboo words were frequently confused with the latter as prerecognition hypotheses.

Goldberg and Fiss (1959) suggested that this "partial information" hypothesis applies to the "discrimination without awareness" literature as well as to the subception data, and found with tachistoscopically exposed geometric figures that partial cues facilitated performance, although when the stimulus was only a blur, the frequency of correct identifications did not exceed chance. They therefore interpreted discrimination without awareness as a laboratory artifact resulting from a difference between the amount of information the subject recognizes he has available and his guess about the identity of the stimulus object.

In a 1960 study of the perception of partial cues, Weiner and Schiller (1960) found that the number of correct identifications of ambiguous figures was directly related to the stimulus intensity

of the identifying cue. They further strengthened the case for the perception of partial cues by demonstrating that the GSR generalization gradient in subliminal perception is related to the structural similarity in contrast to the semantic relatedness of verbal stimuli and that the amount of recognition that occurs in subliminal perception relates to the amount of peripheral stimulation the subject can perceive. Their finding dealing with the GSR gradient contrasts with Spence's (1961) report that subliminal words may facilitate the conscious discrimination of synonyms as well as stimulus words themselves and points out that the extent and type of generalization also depend, among other things, upon the response measures used.

*Subliminal Perception and Phenomena of Learning.* Identifying the role of partial cues in subliminal perception prompts an examination of these phenomena in terms of learning variables. Boardman (1957) reported that the number of correct elements in a series of prerecognition hypotheses increases over trials. This indicates a progressive lowering of the threshold for elements with practice and dictates the need for a methodological control to insure that stimuli are in fact subliminal when they are presented. Meanwhile, Dulany (1957) suggested that the complementary modes of perceptual response—vigilance and defense—can be reconciled as the result of avoidance learning. Perceptual defense is held to be a learned response produced by the punishment of a critical response while competing responses are instrumental to avoidance of punishment and consequent anxiety reduction. Vigilance results from the opposite set of circumstances.

In an experiment to test his notion, he had subjects rate recognizability of simple geometric figures; he shocked some whenever a critical stimulus ap-

peared clearest and others whenever any noncritical figure was presented. In posttraining recognizability tests a predominant number of "defense" subjects displayed a reduction in clarity of the critical figures while a comparable number of "vigilance" subjects showed the opposite tendency. Similarly, Banks and Walters (1959) found that verbal reinforcement of taboo words in a guessing game resulted in reduced recognition thresholds for this class of stimuli in recognition tests. In contrast, Goldstein and Himmelfarb (1962) reported that tachistoscopically presented, emotionally toned words were less frequently recognized than neutral words, and that knowledge of results on the first 25% of the trials had no effect on performance.

Meanwhile, Lowenfeld, Rubenfeld, and Guthrie (1956) reported that the GSR response identified with subception in the Lazarus and McCleary (1951) experiment cannot be inhibited by verbal instructions as long as the reinforced stimuli remain below the recognition threshold. In a sequel to this study of verbal inhibition, Rubenfeld, Lowenfeld, and Guthrie (1956) examined the possibility of stimulus generalization in subception. Subception effects were obtained and stimulus generalization of GSR responses occurred to a greater extent when figures were confused in identification than when they were accurately discriminated.

*Set and the Recognition Threshold.* It is amply clear that recognition thresholds are not solely a function of the properties—either physical or psychological—of the particular stimulus giving rise to the measurement. As Boardman (1957) reported, the accuracy of prerecognition hypotheses on individual trials increases as a test-sequence proceeds, so that what may be initially a blur during tachistoscopic presentation

early in an experimental series may be quite clear at the same speed later on. Thus test attitude, identified with task familiarity, may be an important consideration in the correct interpretation of defense and other subliminal phenomena. The effects of attitudinal factors are seen in other ways. If a class of words is identified by the subject during the test series, expectations resulting from this are likely to produce confusion in the prerecognition hypotheses. In this connection, the presence of offensive words in Boardman's experiment led to the use of taboo-word hypotheses with neutral stimuli of similar physical appearance. Here one has an elevation of threshold quite the opposite of that found in defense experiments, where the recognition of a taboo word is withheld by the persistence of neutral-word hypotheses.

The directive effect of sets where the result is one of enhancing perceptual acuity is also seen in an experiment by Hoisington and Spencer (1958). Lists of words were presented at a distance such that no words on any list could be correctly identified. However, when the subject was given one word from each list and asked to locate its position in the list it fairly "popped" into clarity. This is suggestive of the role of partial cues in recognition. When the subject knows what to look for he has less difficulty in organizing his perceptual field so as to recognize the critical stimuli.

Before such effects can be unequivocally identified as effects upon perception their relation to response bias must be clarified. As noted earlier, a number of students of the defense phenomenon (e.g., Eriksen, 1958; Goldiamond, 1958) have proposed to identify this with response bias rather than with perceptual inhibition. In their view elevated recognition thresholds for taboo words result from the smaller likelihood that subjects

in the formal laboratory situation will entertain prerecognition hypotheses for the category of taboo words. Goldstein (1962) and Goldstein, Himmelfarb, and Feder (1962) have reported several experiments that are directed toward segregating the roles of the perceptual and response variables. In one, subjects were required to identify flashed stimuli (either nonsense syllables or words) from a word list of both neutral and anxiety-arousing words. Subjects given word stimuli displayed a defense effect, correctly identifying fewer emotionally toned than neutral words. However, subjects tested with nonsense syllables showed the same defense effect, prompting the conclusion that defense may result from response bias. In the other experiment, designed to eliminate response bias, no evidence of defense was apparent. While this result may seem to support the identification of defense with response bias, it may also be attributable to a blocking of the anxiety-reducing function of defense by the forced-choice experimental method used.

The need to differentiate between perception and response bias is nicely illustrated by an experiment by Goldiamond and Hawkins (1958). These experimenters required subjects to learn nonsense syllables and then take a visual recognition test. During this latter test no words were in fact flashed on the screen, but "recognition thresholds" were found to be related to the training frequency for the nonsense syllables.

A recent experiment of Phares (1962), reminiscent of the Lazarus and McCleary (1951) study, relates to the anxiety-reduction hypothesis of defense. When pretest avoidance learning was possible because of a systematic relation between the presentation of shock and certain nonsense syllables, recognition thresholds for these syllables were lowered. In contrast, when avoidance learning was

thwarted by a random relationship between the presentation of syllables and shock, recognition thresholds showed no significant drop. Furthermore, this between-conditions difference in recognition threshold held equally well for shocked and nonshocked syllables, suggesting a generalization of the anxiety-reduction effect. Here, however, defense would appear to be identified with the failure of anxiety reduction.

*Affect and Personality.* The early experiments on defense and subception were directed toward exploring the role of the affective or affect related properties of stimuli upon perception. The articulation of this line of research seems frequently to have been interfered with by methodological issues and the intrusion of other variables. In one of the few studies in the last few years concerned directly with affect, Lowenfeld (1961) describes experimental analogues of repression, subception, and defense resulting from negative affect induced into verbal stimuli through applying shock to corresponding response stimuli. This latter technique extends that of Lazarus and McCleary (1951), in that it probably simulates the operations by which affective tone is induced outside the laboratory.

Related to this interest in the mechanisms by which affect relates to the subliminal phenomena are considerations of individual differences and personality variables. Pustell (1957) found striking sex differences in response to threat stimuli. Males showed a predominant tendency to display vigilance in the face of threat stimuli, while females displayed a less marked tendency toward defense. To explain this Pustell fell back upon a suggestion of Dollard and Miller that when stimulation is mild it functions as a cue whereas when it is more intense it acquires drive properties. By assuming that his female subjects experienced

more intense anxiety during testing, he saw defense as a means of reducing an unpleasant drive. Vigilance is the male response to what was for them a less severe stress-cue. W. P. Brown (1961) experimentally indicated that sex differences in the defense response may be quite complex, and Van de Castle (1960), using the binocular-rivalry response, reported that sensitizers and defenders, as identified by performance on the Welsh *A* and *R* scales, differed in the readiness with which they perceived aggressive words when pitted against neutral words. These two classes of subjects also differed in their responses to the Rorschach.

*Subliminal Stimulation and Perception and Judgment.* The methodological problems that face the experimenter when using subliminal stimulation as an independent variable are essentially those that have been described for the literature on subliminal perception. Thus, many experiments have been arrested at the point of evaluating whether or not such stimulation can have an effect upon behavior. Meanwhile, a variety of behavioral phenomena has been examined as dependent variables. Goldstein (1960) selected six common and similar geometrical illusions with distorting backgrounds as subliminal stimuli and required subjects to identify the shape of the focal stimuli from an ordered set of nondistorted and distorted figures. Judgments overall tended toward distortion although such effects were not always evident for specific illusions on selected trials. Here the criterion of subliminal stimulation was that the subject displayed ignorance of the presence of the distorting background on posttest inquiry. Similarly, Walters, Banks, and Ryder (1959) reported that the subliminal flashing of taboo words prior to the presentation of neutral words resulted in distortion

or misperception of words that were themselves in no way taboo.

Subliminal stimuli also influence processes that are more clearly judgmental than perceptual. For example, Smith, Spence, and Klein (1959) indicated that flashing the words HAPPY or ANGRY prior to the suprathreshold presentation of a relatively expressionless face resulted in the verbal descriptions of the face as either pleasant or unpleasant. In a related study, Klein, Spence, Holt, and Gourevitch (1958) presented drawings of sexually ambiguous human figures preceded by subliminally flashed drawings of male or female genitals or male or female symbols. Here, subjects were required to draw the figures they had been presented as well as make adjective check list ratings concerned with the masculinity-femininity of the figures. Subliminal presentation of genital drawings influenced both check list and drawing responses toward the sex indicated by the flashed figure, but the symbolic figures influenced only the masculinity or femininity of the drawings.

*Subliminal Stimulation and Guessing and Choice Behavior.* Closely related to the operations of detection and discrimination are the functions of guessing and choice. The literature on "discrimination without awareness" has essentially consisted of demonstrations of nonrandom series of responses to stimuli below threshold, and, in the recent past, there has been a practical interest in the possibility that weak or unobtrusive stimulation would prompt certain kinds of choice behavior such as buying refreshments at the movies or selecting certain brands of merchandise above others. This latter interest has given rise to a number of experiments devised to test the claims of subliminal advertising. A study by Champion and Turner (1959) is representative. During the showing of a film on sales administration, a drawing

of a spoonful of rice with the caption, "Wonder Rice," was flashed periodically. Control subjects were shown a random array of straight lines on a similar schedule. After the film, the slide of the spoon was shown and the subjects were asked to indicate if they had ever seen it as advertising and what brand they thought it represented. The results of the experiment on both counts were negative.

In order to achieve laboratory control better, Calvin and Dollenmayer (1959) used simplified stimulus materials and turned the task into a guessing game by presenting the subject on successive trials with instructions, "choose right" or "choose left," flashed subliminally on the screen. Again, as long as the instructions remained subliminal, they failed to influence choice behavior. Negative results were also obtained by an experiment by Vernon and Badger (1959).

In contrast, Schiff (1961) reported positive results when subjects were directly engaged in a guessing experiment in which the experimental group received subliminal flashes of geometrical figures to be guessed and the control group got blank flashes. Although all presentations of the figures were below each experimental subject's lowest measured absolute threshold, this experimental precaution does not entirely compensate for the possibility that recognition thresholds drop with repeated testing and that the threshold of verbal identification may not be coextensive with perception. Sharp (1959) did a subliminal stimulation experiment under the guise of testing subject-matter knowledge. Either correct or incorrect answers were presented subliminally as the experimental condition. While the cues were reported to influence performance significantly, it was also found that 60% of the subjects learned to detect their presence

consciously. Furthermore, Giddan and Eriksen (1959) have demonstrated that unobtrusive patterns of reinforcement can induce systematic response biases in discrimination behavior just as they can in verbal conditioning.

Still another study is that of Spence and Bressler (1962) who flashed the word HOUSE, then a blurred array of nonsense words. The subject was then presented a list of frequent and infrequent associates to HOUSE, as well as a set of control words, and asked if any of them were in the blurred array. Different groups received the key word subliminally, near threshold, and supraliminally. Results indicated that reaction time was negatively correlated with associative strength for the subthreshold condition, but failed to relate to associative strength for the other two conditions.

More complex tasks have been used to examine the effect of subliminal stimulation upon cognitive functions. Kolers (1957) used metacontrast to present subliminal cues for the solution of two types of perceptual problems. In one, three sets of geometrical figures were flashed simultaneously and the subject was to identify a figure common to all sets. (In fact, there were two figures common to all sets and the subliminal cue consisted of one of these.) In the other, the subject had to identify in one set of embedded figures a figure common to all members of another. (Here the cue was the embedded figure.) In the case of the first task, experimental subjects identified the subliminal figure as correct more frequently than did the controls. In the second task, experimental subjects solved more problems correctly than did the controls. In a later study Gerard (1960) replicated the Kolers experiment with the added condition of subliminal presentation of incorrect solutions. His results contra-

dict those of Kolers and defy ready explanation, for both experimental groups (correct and incorrect cues) made fewer correct responses than did the control subjects who received no cues.

The literature which has appeared since the McConnell (1958) review confirms its conclusion. The influences of subliminal stimulation upon preference and choice, if they occur at all, are highly subtle and the possibility that they could constitute an effective means of controlling consumer behavior or political opinion is highly unlikely. Similarly, the technique of incidental reinforcement of the verbal conditioning studies appears less potent than the early studies of this phenomenon have suggested. For example, Kriekhaus and Eriksen (1960) studied the effect of verbal reinforcement in the choice of verbal alternatives. While an effect was obtained in subjects unaware of the relationship between choice and reinforcement, it was much less than that found in subjects aware of this relationship. Furthermore, the effect produced failed to generalize to any other aspect of the test stimuli.

*Subthreshold Stimulation and Motivation or Affect.* While it would appear that subthreshold stimulation has at best an unreliable effect upon choice as a generic class of behavior, it is conceivable that it may influence motivation or affect, variables related to choice behavior. Two recent experiments bear on this point. Byrne (1959) demonstrated that experimental subjects receiving experimental flashes of the word, "beef," during an instructional film rated themselves as hungrier than did control subjects. Subsequent to the film, the experimental subjects showed no tendency to make verbal references related to the subliminal stimulus in sentence completion or word association tests and no greater tendency to prefer objects re-

lated to the key word in a preference test. Hence, subliminal stimulation may influence drive arousal if not recognition thresholds.

On the other hand, subjects in an experiment by Goldstein and Davis (1961) who were unaware of the words, "good" and "bad," flashed beneath photographs failed to be influenced by them whereas subjects who saw the words were affected.

*Subthreshold Stimulation and Elaborative Thought.* Three recent experiments bear on this problem. All provide evidence supporting a subliminal effect. In a study in which subjects gave responses to TAT cards while negatively or positively toned verbal cues were tachistoscopically presented, Goldstein and Barthol (1960) found that when the test cards were sufficiently ambiguous, a significant influence of the words was apparent. Zuckerman (1960), in an experiment also involving TAT cards, gave subliminal instructions to either write more or less. While control subjects showed a steady increase in amount produced for consecutive cards, experimental subjects showed a decrease with a change in instructions from more to less. Pine (1960) had subjects freely elaborate stories and found that the content of the productions could be influenced by a passage heard through the wall of an adjacent room. This effect, however, was less pronounced than when the passage had been read by the subject himself.

*Subliminal Stimulation and Autistic Responses.* Two studies are concerned with this relationship. In a 1958 experiment, Shevrin and Luborsky reported a confirmation of the Poetzl phenomenon, the tendency for tachistoscopically presented material to appear in dreams when not available in intentional recall.

These experimenters found a greater number of conceptually similar items in dream material than in earlier attempts at recall. Johnson and Eriksen (1961), questioning the appropriateness of the criterion of conceptual similarity as well as the use of a cohesive scene as stimulus material, did a similar experiment with what they regarded as stricter experimental controls and failed to confirm the Shevrin and Luborsky results.

*In Summary.* A broad glance across these recent data supports the conclusion that subliminal perceptual effects are real effects, though so sensitive in production as not to be seriously considered as a technique to be exploited in the interest of consumer sales or other practical goals. It is also clear that certain results are artifacts of procedure. Meanwhile, part of the problem involves a semantic—if not, indeed, a metapsychological—issue. It is interesting to see the classical question of levels of consciousness reappear in this context. In the evolution of concepts, movement of thought may be said to transcribe a helixlike path, conceptual problems being brought full circle but at a higher level. Progress in the psychologist's dealing with levels of awareness is seen in his application of the device of converging operations to this problem. But, as we have just noted, the problem is in part a semantic problem. Thresholds are identified in terms of arbitrarily specified operations—for example, an intensity reported 50% of the time—and it is perfectly proper to define as subthreshold events that occur reliably but that do not meet these definitional criteria. Meanwhile, what is most important for an understanding of behavior is not that these effects are unconscious, but that stimuli, subtle beyond perception in any ordinary sense, can influence a variety of response systems.

# SUBLIMINAL STIMULATION AND JUDGED MAGNITUDE

There are, then, three major types of effect associated with subliminal stimulation. One involves the evocation of responses that are essentially equivalent to those produced by supraliminal stimuli. For example, a local irritant on the surface of the skin may produce scratching while a subject is preoccupied with another activity and unaware of his scratching. The second is characterized by the failure at appearance of responses which one otherwise might expect to appear. The perceptual defense data fall into this category. The third type is marked by the induction of a change or distortion in the perceptual responses to a supraliminal stimulus. The final section of this paper is devoted to a review of experiments of my own dealing with this last effect.

For 10 years the present writer has been interested in exploring the psychophysical model developed by Helson (1959) in connection with his adaptation-level theory. Within the context of this theory, the organism is viewed as a system for taking inputs from various sources—stimuli, their background, and past experience—and combining these to constitute internal norms or standards which become the reference values for subsequent judgments. Thus the brightness of a light, for example, is not simply a matter of the physical intensity of the light stimulus, but depends also upon the brightness of the background on which it appears and the impression of average brightness evolved by the judge from his past experience with lights. Preoccupation with the organism as a pooling norm-deriving system forces interest in the problem of relevance. By relevance is meant the criteria that differentiate the inputs which will be pooled from those which are ineffective in the

derivation of particular subjective judgmental norms. Interest in the general question of relevance has led us to a consideration of many specific problems. The pooling model, for example, provides a new view of behavioral constancy. Bevan and Darby (1955) were able to show that different frequency-intensity patterns of input can, within certain limits, yield the same indifference point for lifted weights. Several experiments in the writer's laboratory have indicated that sense modality is not a limiting condition of relevance. Bevan and Pritchard (1963b) have reported that visual anchors can be used to influence the judged loudness of tones and Behar and Bevan (1961), in an extended series on judged duration, found that they could not only induce, but they could precisely predict, heteromodal anchor effects.

In the present experiments the writer's interest in relevance was extended to include the absolute threshold as a possible limiting condition for pooling. Not only did Fechner make the magnitude of the present stimulus process the sole determinant of the corresponding sensory attribute, but also, in his formulation of the psychophysical law,  $S = k \log R$ , he assumed that the absolute threshold is the origin of the sensory scale, a fixed and limiting value. Like Fechner, but for different reasons, Wundt (Boring, 1929, p. 332), and later Titchener (Bevan, 1958, p. 37), held the absolute threshold to be the limiting value of all psychologically relevant stimuli. This position has been retained by Stevens (1936) and other psychophysicists in the classical tradition, but adaptation-level theorists (Michels & Helson, 1949) have shown that Fechner's law can be derived on rational grounds with prevailing adaptation level as the origin of this psychophysical function. The advent of a modern functionally oriented psycho-

physics, with its emphasis on the pooling of background and residual stimuli with focal stimuli, has made the possibility of subliminal effects highly plausible. The experiments to be reviewed now sought to determine if subliminal stimuli influence a subject's responsiveness to supraliminal stimuli. The addition of a subliminal stimulus to a supraliminal series fits the paradigm of the anchor experiment. Should the subliminal stimulus behave like a supraliminal anchor below the series, it may be expected to produce an elevation in judged intensity of the judgments of the suprathreshold series members. Furthermore, this change should be greatest for the stimuli at the lower end of the series; that is, closest to the subliminal anchor.

The experiments to be described in the following paragraphs have two methodological features in common. In no instance has the subliminal stimulus evoked a positive response nor was its inclusion in the order of presentation known. Subjects who responded to one or more presentations of the subliminal stimulus or who, upon questioning after test, indicated cognizance of its presence were eliminated from the experiments.

*Shock Experiments.* The first experiment was a pilot study. A small number of subjects made category judgments of a series of mild electric shocks under two conditions, the experimental sequence during which a subliminal shock was presented, without the subject's knowledge, at the midpoint of the presentation interval and the control sequence, without the subliminal anchor. When judged intensity of the series was plotted against its physical intensities, the values for three stimuli under the experimental condition lay, as expected, above corresponding control values. However, the average differences between conditions was small and failed statistical significance. Meanwhile, there was

a significant stimuli-by-conditions interaction, indicating a difference in the slopes of the functions and suggesting a subliminal anchor effect. Further data analysis revealed that subjects, regardless of the order of the conditions, tended to give more intense judgments for the first than for the second condition. Such a temporal effect might be expected to mask the effects of the experimental treatment. Order effects are a common bugaboo in psychophysical experiments when counterbalancing is used. Therefore, the experiment was repeated with an independent group for each condition.

In Experiment II (Black & Bevan, 1960), GSR measurements were made for all series stimuli and the subliminal anchor or the midpoint of the presentation interval as appropriate, with the hope of registering the arousing effects of the subliminal stimulus although its presence went undetected. Statistical comparison of the experimental and control functions revealed that the judgments of the experimental subjects were consistently more intense than those for the controls. Furthermore, the greatest elevation of the experimental curve occurred in the low end of the series, that is, nearest the subliminal stimulus on the intensity dimension. Thus, it would appear that the subliminal stimulus behaved like an anchor below the series. However, there was no evidence of more than the chance occurrence of GSRs with the presentation of the subliminal anchor, suggesting the possibility that there are circumstances for which behavioral measurements are more sensitive indices of stimulation than are available physiological measures. It also suggests that the brain-stem mechanisms of arousal may differ from those associated with anxiety and perceptual defense.

Goldstone, Goldfarb, Strong, and Russell (1962) have recently repeated this last experiment with a change in proce-

ture designed to enhance the effectiveness of the subliminal anchor. Instead of interpolating the anchor at the midpoint of the presentation interval between series stimuli, this stimulus was presented just prior to each series stimulus after the first. Again, the average judged intensity for the experimental subjects was consistently higher than that for the controls, and the mean difference between groups was statistically significant.

*Generality of the Shock Results.* Being satisfied that subliminal shocks do, indeed, influence the apparent magnitude of series shocks above threshold, the next question of concern was whether or not this effect is confined to electrical stimulation or whether or not it is more general in nature.

Accordingly, Experiment II was replicated with loudness substituted for shock as the dimension to be judged (Bevan & Pritchard, 1963a). Because the test room was poorly insulated for sound, a 70-decibel sound pressure level, white-noise sound screen was set up. Comparison of the experimental and control data revealed both highly significant between-group slopes and between-group means sources of variance. Curves for the two conditions were most widely separated, as expected, at the low end of the series. But, unexpectedly, the relative positions of the curves were reversed, thus contradicting both the prediction and the results of the early experiments. Two possibilities for reconciliation presented themselves. One was that, for the particular set of conditions selected for the experiment, the effect was one of assimilation rather than contrast, the anchor attracting rather than repelling series judgments. The other possibility was that the impressive sound screen, while it kept extraneous noises from producing anchor shifts, also itself served as a potent anchor above the series and, perhaps, interacting with

the anchor designate, produced the unexpected results.

Therefore, the experiment was repeated in a specially constructed sound-insulated laboratory, the experimental subject, in addition, wearing highly effective ear muffs to screen out any external sounds that might get through room insulation. The data from this experiment were ambiguous. The judgments of the experimental group fell above those of the control for the three highest series magnitudes, coincided on the next to the lowest magnitude, and was below the control value on the weakest of the series stimuli. However, the overall difference between the average judgments for control and experimental conditions was clearly significant. Thus, it would appear that if there were an anchor effect, it was a weak one. It would also appear that the possibility of an assimilation rather than contrast effect for the stimuli nearest the anchor was plausible.

Ambiguity of these results bred discontent. Hence, yet another experiment was performed, the purpose of which was twofold: to seek a more effective set of conditions for the production of the subliminal anchor effect and, as part of this, to explore the result of changing the temporal position of the anchor within the presentation interval. As was indicated earlier, Goldstone, Goldfarb, Strong, and Russell (1962) reported that while an attempt to reproduce the Black-Bevan results with the anchor at the midpoint of the interval was unsuccessful, it was possible to do so by moving the anchor back in time so that it preceded the next series stimulus by a shorter interval. Following this line of reasoning, it was expected that the ambiguity of the last results might be clarified by using several anchor-series stimulus intervals. It was hypothesized that the anchor effect might be enhanced by increasing the scalar distance be-

tween anchor and series. At the same time, this meant, of course, using a weaker physical stimulus as anchor. The control group of the last experiment was used as the control for the present study. In addition, three anchor groups were tested. In one, the subliminal anchor was presented just prior to each series stimulus to be judged, in the second at the midpoint of the presentation interval, and in the third just after each series stimulus.

The results are rather interesting. In every case, the experimental group made reliably more intense judgments than the control. This is the contrast effect that was sought. It is consistent with the original results of Black and Bevan (1960). Furthermore, the position of the anchor within the interstimulus interval seems not to have had any effect upon the magnitude of the anchor effect. This, of course, does not mean that interpolation time is not an important general consideration, but only that under the particular experimental conditions of this last experiment, it failed to operate as a significant variable.

An experiment similar to these auditory experiments has been reported by Boardman and Goldstone (1962). Their experimental task involved the judgment of the size of discs. Using the visual masking technique with a four-field electronic tachistoscope, one group of subjects received the psychophysical series along with a subliminal anchor smaller than any of the series members, while another was presented the series and a subliminal anchor larger than the series members in size. Each subject served as his own control. Although the differences between control and experimental conditions were small and were not significant for all series magnitudes, the results were in line with expectation. For the group receiving the small anchor, the experimental values plotted above the control values and the greatest differ-

ence was at the low end of the series. For the group receiving the anchor above the series, the control values plotted above those of the experimental and the greatest spread was at the upper end of the series.

*Some Implications of the Data on Magnitude Judgment.* There are now enough data available to prompt the conclusion that the subliminal anchor effect is a real one. Thus one is faced with the responsibility of reflecting upon its significance. These results, it appears, have implications for two or three areas of inquiry. Although more experimentation is needed, if the results on GSR in the electric shock experiment are valid, they suggest the methodological rule of thumb that there are situations in which behavioral indices are more sensitive measures of mediating events than are the usual physiological measures. The fact that not a single GSR occurred to the presentation of the subliminal stimulus may imply that the brain-stem mechanisms associated with arousal and vigilance are distinct from those related to anxiety and fear. Perhaps, when more is learned of the properties of the reticular activity system (Jasper, 1958), these interpretations of the present results will receive experimental support.

The question concerning relevance which prompted these experiments is, of course, answered. The absolute threshold cannot be applied as a criterion in the identification of stimuli relevant for pooling. This has implications not only for adaptation-level theory, but for psychophysics as well, for it draws suspicion to the long-standing assumption that the traditionally defined absolute threshold is the limiting value for all sensory dimensions. This suggests, in turn, that sensory scales, even those constructed with the so-called ratio methods, are not absolute but relative, since the origin of the scale is not a true zero.

Finally, it may be suggested that the results of these experiments may have some value for an understanding of how unconscious mechanisms exert their influence upon behavior. The nature of the unconscious has been most thoroughly articulated within the context of psychoanalytic theory. From this point of view, thought goes on at two levels: primary thought (Freud, 1946), identified with the unconscious, is characterized as essentially dynamic, directed by instinctual urges, amoral, and governed by a set of laws qualitatively different from the conventional logic of secondary, or conscious goal-directed thinking (Peck, 1950). Where primary thought is directed toward the fulfillment of the pleasure principle, secondary thought serves the reality principle. The essential nature of the relationship between these two levels is said to be one of conflict, conscious thought sitting as censor of the unconscious (Jones, 1948). The particular mechanism by which unconscious stimuli (and in a formal sense, instinctive drives are stimuli) exert their influence upon overt behavior is, however, unclear. Meanwhile, the general impression one gets from the Freudian literature is that the relationship is direct, automatic, and, in the usual sense of the term, irrational. The data of these experiments join other data in support of the view that at least one device by which unconscious variables influence behavior is through their role in structuring the behavioral environment within which the behavior they influence occurs. Here their role is similar to that of supraliminal variables and the behavior, far from being irrational, is not only rational but appropriate to the circumstances as the behaving individual perceives them.

# REFERENCES

- ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, 54, 383-405.
- BAKER, L. E. The influence of subliminal stimuli upon verbal behavior. *J. exp. Psychol.*, 1937, 20, 84-100.
- BANKS, R. K., & WALTERS, R. H. Prior reinforcement as a determinant of recognition thresholds. *Percept. mot. Skills*, 1959, 9, 51-54.
- BEHAR, I., & BEVAN, W. The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *Amer. J. Psychol.*, 1961, 74, 17-26.
- BEVAN, W. Perception: Evolution of a concept. *Psychol. Rev.*, 1958, 65, 34-55.
- BEVAN, W., & DARBY, C. L. Patterns of experience and the constancy of an indifference point for perceived weight. *Amer. J. Psychol.*, 1955, 68, 575-584.
- BEVAN, W., & PRITCHARD, JOAN FAYE. Effect of "subliminal" tones upon the judgment of loudness. *J. exp. Psychol.*, 1963, 66, 23-29. (a)
- BEVAN, W., & PRITCHARD, JOAN FAYE. The effect of visual intensities upon the judgment of loudness. *Amer. J. Psychol.*, 1963, in press. (b)
- BLACK, R. W., & BEVAN, W. The effect of subliminal shock upon the judged intensity of weak shock. *Amer. J. Psychol.*, 1960, 73, 262-267.
- BOARDMAN, W. K. Utilization of word structure in prerecognition responses. *J. Pers.*, 1957, 25, 672-685.
- BOARDMAN, W. K., & GOLDSTONE, S. Effects of subliminal anchors upon judgments of size. *Percept. mot. Skills*, 1962, 14, 475-482.
- BORING, E. G. *A history of experimental psychology*. New York: Appleton Century, 1929.
- BROWN, J. S. *The motivation of behavior*. New York: McGraw-Hill, 1961. Ch. 8.
- BROWN, W. P. Conceptions of perceptual defense. *Brit. J. Psychol.*, 1961, Monogr. Suppl. No. 35.
- BYRNE, D. The effect of a subliminal food stimulus on verbal responses. *J. appl. Psychol.*, 1959, 43, 249-252.
- CALVIN, A. D., & DOLLENMAYER, K. S. Subliminal perception: Some negative findings. *J. appl. Psychol.*, 1959, 43, 187-188.
- CARPENTER, B., WEINER, M., & CARPENTER, JANETTE T. Predictability of perceptual defense behavior. *J. abnorm. soc. Psychol.*, 1956, 52, 380-383.
- CHAMPION, J. M., & TURNER, W. W. An experimental investigation of subliminal perception. *J. appl. Psychol.*, 1959, 43, 382-384.
- DAVIS, J. M. Personality, perceptual defense, and stereoscopic perception. *J. abnorm. soc. Psychol.*, 1959, 58, 398-402.
- DAVIS, K. G., & HESS, H. F. The effectiveness

- of concepts at various levels of awareness. *J. exp. Psychol.*, 1962, 63, 62-67.
- DAY, W. F. Serial non-randomness in auditory difference thresholds as a function of inter-stimulus interval. *Amer. J. Psychol.*, 1956, 69, 387-384.
- DIXON, N. F. Apparent changes in the visual threshold as a function of subliminal stimulation. *Quart. J. exp. Psychol.*, 1958, 10, 211-219. (a)
- DIXON, N. F. The effect of subliminal stimulation upon autonomic and verbal behavior. *J. abnorm. soc. Psychol.*, 1958, 57, 29-36. (b)
- DULANY, D. E., JR. Avoidance learning of perceptual defense and vigilance. *J. abnorm. soc. Psychol.*, 1957, 55, 333-338.
- EDWARDS, A. E. Subliminal tachistoscopic perception as a function of threshold method. *J. Psychol.*, 1960, 50, 139-144.
- ERIKSEN, C. W. Defense against ego threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, 47, 230-235.
- ERIKSEN, C. W. An experimental analysis of subception. *Amer. J. Psychol.*, 1956, 69, 621-634. (a)
- ERIKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, 63, 74-80. (b)
- ERIKSEN, C. W. Unconscious processes. In M. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958. Pp. 169-226.
- ERIKSEN, C. W. (Ed.) *Behavior and awareness: A symposium of research and interpretation*. Durham: Duke Univer. Press, 1962.
- ERIKSEN, C. W., AZUMA, H., & HICKS, R. Verbal discrimination of pleasant and unpleasant stimuli prior to specific identification. *J. abnorm. soc. Psychol.*, 1959, 59, 114-119.
- FREUD, S. Formulations regarding the two principles in mental functioning. In *Collected papers*. Vol. 4. London, England: Hogarth Press, 1946. Pp. 13-21.
- FUHRER, M. J., & ERIKSEN, C. W. The unconscious perception of the meaning of verbal stimuli. *J. abnorm. soc. Psychol.*, 1960, 61, 432-439.
- GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, 63, 149-159.
- GERARD, E. O. Subliminal stimulation in problem-solving. *Amer. J. Psychol.*, 1960, 73, 121-126.
- GIDDAN, N. S., & ERIKSEN, C. W. Generalization of response biases acquired with and without verbal awareness. *J. Pers.*, 1959, 27, 104-115.
- GILCHRIST, J. C., LUDEMAN, J. F., & LYSACK, W. Values as determinants of word recognition thresholds. *J. abnorm. soc. Psychol.*, 1954, 49, 423-426.
- GOLDBERG, F. R., & FISS, H. Partial cues and the phenomenon of discrimination without awareness. *Percept. mot. Skills*, 1959, 9, 243-251.
- GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, 55, 373-411.
- GOLDIAMOND, I., & HAWKINS, W. F. Vexier-versuch: The log relationship between word frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, 56, 475-463.
- GOLDSTEIN, M. E. Subliminal perception with optical illusions. *J. gen. Psychol.*, 1960, 62, 89-101.
- GOLDSTEIN, M. J. A test of the response probability theory of perceptual defense. *J. exp. Psychol.*, 1962, 63, 23-28.
- GOLDSTEIN, M. J., & BARTHOL, R. P. Fantasy responses to subliminal stimuli. *J. abnorm. soc. Psychol.*, 1960, 60, 22-26.
- GOLDSTEIN, M. J., & DAVIS, DIANE. The impact of stimuli registering outside of awareness upon personal preference. *J. Pers.*, 1961, 29, 247-257.
- GOLDSTEIN, M. J., & HIMMELFARB, S. Z. The effects of providing knowledge of results upon the perceptual defense effect. *J. abnorm. soc. Psychol.*, 1962, 64, 143-147.
- GOLDSTEIN, M. J., HIMMELFARB, S. Z., & FEDER, WALDA. A further study of the relationship between response bias and perceptual defense. *J. abnorm. soc. Psychol.*, 1962, 64, 56-62.
- GOLDSTONE, S., GOLDFARB, JOYCE, STRONG, J., & RUSSELL, S. Replication note: The effect of subliminal shock upon the judged intensity of weak shock. *Percept. mot. Skills*, 1962, 14, 222.
- GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, 68, 409-416.
- HALL, J. F. *Psychology of motivation*. Chicago: Lippincott, 1961. Ch. 12.
- HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. I. *Sensory, perceptual and physiological formulations*. New York: McGraw-Hill, 1959. Pp. 565-621.
- HOISINGTON, L. D., & SPENCER, CAROL. Specific set and the perception of "subliminal" material. *Amer. J. Psychol.*, 1958, 71, 263-269.
- HOWES, D., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, 57, 235-240.

- JASPER, H. Reticular-cortical systems and theories of integrative action of the brain. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958. Pp. 37-63.
- JENKINS, N. Affective processes in perception. *Psychol. Bull.*, 1957, 54, 100-127.
- JOHNSON, H., & ERIKSEN, C. W. Preconscious perception: A re-examination of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1961, 62, 497-503.
- JONES, E. *What is psychoanalysis?* New York: International Universities Press, 1948.
- KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNAH. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, 57, 255-266.
- KOLERS, P. A. Subliminal stimulation in problem solving. *Amer. J. Psychol.*, 1957, 70, 437-441.
- KRIECKHAUS, E. E., & ERIKSEN, C. W. A study of awareness and its effects on learning and generalization. *J. Pers.*, 1960, 28, 503-517.
- LAZARUS, R. S., & MCCLEARY, R. A. Automatic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, 58, 113-122.
- LOWENFELD, J. Negative affect as a causal factor in the occurrence of repression, subception, and perceptual defense. *J. Pers.*, 1961, 29, 54-63.
- LOWENFELD, J., RUBENFELD, S., & GUTHRIE, G. M. Verbal inhibition in subception. *J. gen. Psychol.*, 1956, 54, 171-176.
- MCCLELLAND, D. C., & LIEBERMAN, A. M. The need for achievement in recognition of need-related words. *J. Pers.*, 1949, 18, 236-251.
- MCCONNELL, J. V., CUTLER, R. L., & MCNEIL, E. B. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, 13, 229-242.
- MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, 56, 244-251.
- MATHEWS, ANNE, & WERTHEIMER, M. A "pure" measure of perceptual defense uncontaminated by response suppression. *J. abnorm. soc. Psychol.*, 1958, 57, 373-376.
- MICHEL, W. C., & HELSON, H. A reformulation of the Fechner law in terms of adaptation-level applied to rating scale data. *Amer. J. Psychol.*, 1949, 62, 355-368.
- PECK, M. W. *The meaning of psychoanalysis*. New York: Permabooks, 1950.
- PHARES, E. J. Perceptual threshold decrements as a function of skill and chance expectancies. *J. Psychol.*, 1962, 53, 399-408.
- PINE, F. Incidental stimulation: A study of preconscious transformations. *J. abnorm. soc. Psychol.*, 1960, 60, 68-75.
- PUSTELL, T. E. The experimental induction of perceptual vigilance and defense. *J. Pers.*, 1957, 25, 425-438.
- RUBENFELD, S., LOWENFELD, J., & GUTHRIE, G. M. Stimulus generalization in subception. *J. gen. Psychol.*, 1956, 54, 177-182.
- SCHIFF, W. The effect of subliminal stimuli on guessing accuracy. *Amer. J. Psychol.*, 1961, 74, 54-60.
- SHARP, H. C. Effect of subliminal cues on test results. *J. appl. Psychol.*, 1959, 43, 369-371.
- SHEVRIN, H., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, 56, 285-294.
- SMITH, G. J. W., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, 59, 167-176.
- SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, 58, 256-270.
- SPENCE, D. P. A new look at vigilance and defense. *J. abnorm. soc. Psychol.*, 1957, 54, 103-108.
- SPENCE, D. P. The multiple effects of subliminal stimuli. *J. Pers.*, 1961, 29, 40-53.
- SPENCE, D. P., & BRESSLER, JOANNA. Subliminal activation of conceptual associates: A study of "rational" preconscious thinking. *J. Pers.*, 1962, 30, 89-105.
- STEVENS, S. S. A scale for the measurement of a psychological magnitude: Loudness. *Psychol. Rev.*, 1936, 43, 405-416.
- TAYLOR, JANET A., ROSENFELDT, DORIS C., & SCHULZ, R. W. The relationship between word frequency and perceptibility with a forced-choice technique. *J. abnorm. soc. Psychol.*, 1961, 62, 491-496.
- VAN DE CASTLE, R. L. Perceptual defense in a binocular-rivalry situation. *J. Pers.*, 1960, 28, 448-462.
- VERNON, J. A., & BADGER, D. H. Subliminal stimulation in human learning. *Amer. J. Psychol.*, 1959, 27, 47-55.
- WALTERS, R. H., BANKS, R. K., & RYDER, R. R. A test of the perceptual defense hypothesis. *J. Pers.*, 1959, 27, 47-55.
- WEINER, M., & SCHILLER, P. H. Subliminal perception or perception of partial cues. *J. abnorm. soc. Psychol.*, 1960, 61, 124-137.
- ZUCKERMAN, M. The effects of subliminal and supraliminal suggestion on verbal productivity. *J. abnorm. soc. Psychol.*, 1960, 60, 404-411.

(Received September 10, 1962)

# NEURAL CONTROL OF FEEDING BEHAVIOR<sup>1</sup>

ANNETTE EHRLICH

*Northwestern University School of Medicine*

Experiments showing that food intake can be dramatically increased (hyperphagia) or even eliminated (aphagia) as a result of brain lesions or of brain stimulation are described. The most widely accepted explanation of how the brain controls feeding behavior, Anand and Brobeck's dual-center theory of hypothalamic regulation, does not account satisfactorily for all the data, but it does account for a great deal. Experiments are reviewed in which food intake was altered by lesions or by stimulation outside of the hypothalamus, and it is suggested that the hypothalamus is part of a larger "feeding system" that also includes parts of the thalamus, temporal lobe, frontal lobe, and midbrain.

Mohr (1840) published what is generally considered to be the first clinical report of pathological obesity in man resulting from a tumor at the base of the brain (cited by Brobeck, 1946). Thereafter, although similar cases appeared in the literature from time to time (see, for example, Paget, 1897), no particular interest was shown in localizing the effective lesion until the early 1900s, when a heated dispute arose as to whether damage to the pituitary or to the hypothalamus was responsible. Numerous reports of experimentally produced hypothalamic obesity in animals appeared in the literature between 1920 and 1940, but most of these (with the possible exception of Bailey & Bremer's work, 1921) were received with skepticism because the operative procedure used made it unlikely that the pituitary had been completely spared. The difficulty was not overcome until the early 1930s, when the method of making discrete electrolytic lesions by means of the stereotaxic instrument came into general use. At that time, several preliminary reports

appeared in which this method was used (Hetherington & Ranson, 1939; Ranson, Fisher, & Ingram, 1938), but the most conclusive demonstration that hypothalamic lesions alone were responsible for obesity was furnished by Hetherington and Ranson (1940)—exactly 100 years after the publication of Mohr's article. Subsequently, the possibility that pituitary damage played even a partial role was ruled out by hypophysectomizing rats before (Hetherington, 1943) and after (Hetherington & Ranson, 1942a) hypothalamic lesions. The results showed that obesity occurred after hypothalamic lesions whether the pituitary was intact or not, and removal of the pituitary in no way altered the hypothalamic symptoms.

Once the pituitary-hypothalamic controversy had been resolved, most investigators interested in the problem of neural control of food intake concentrated their efforts on elaborating the hypothalamic syndrome, and it is only quite recently that systematic attempts have been made to explore the effects on feeding behavior of other types of brain injury. This review will begin, therefore, with a description of the effects of hypothalamic injury, and the effects of extrahypothalamic lesions will be described in a separate section.

<sup>1</sup>This research was supported by Grant Number 1771 from the Illinois State Psychiatric Research and Training Authority and by Grant Number M-2455 from the United States Public Health Service.

EFFECTS OF HYPOTHALAMIC LESIONS  
ON HUNGER*Hyperphagia*

The hypothalamic lesions made by Hetherington and Ranson (1940) in their pioneer experiment were large, but in a further series of experiments, Hetherington showed that obesity could be produced by destroying only a small hypothalamic area—the ventromedial nucleus and the area immediately adjacent to it (Hetherington, 1941, 1944; Hetherington & Ranson, 1942b). This finding was quickly confirmed by other investigators, and it was shown to be true for several animal species—rats (Tepperman, Brobeck, & Long, 1941), cats (Wheatley, 1944), and monkeys (Brooks, Lambert, & Bard, 1942). In mice, obesity following ventromedial hypothalamic lesions has been demonstrated by two different methods—electrolytic lesions (Mayer, French, Zighera, & Barnett, 1955) and goldthioglucose poisoning (Marshall, Barnett, & Mayer, 1955).

It was fairly soon established that the cause of hypothalamic obesity was overeating or hyperphagia. Hetherington (1941) was inclined at first to attribute the weight gain to underactivity and metabolic disturbances, but paired feeding experiments with normal controls as well as measurements of food intake, metabolic rate, and activity level showed that the changes in activity and metabolism were minor by comparison with the enormous increase in food intake (Brobeck, 1946; Brobeck, Tepperman, & Long, 1943; Brooks, 1946; Brooks, Marine, & Lambert, 1946; Tepperman et al., 1941). With electrolytic lesions, the hyperphagia is apparent as soon as the animal comes out of the anesthetic (Brobeck et al., 1943; Brooks, Lockwood, & Wiggins, 1946), and it continues for a period of several months (dynamic

phase) during which the animal may double its original weight. Eventually, a plateau is reached (static phase) and the food consumption tapers off to normal. Obese animals can be dieted down to a normal weight, but they again become obese when fed ad lib. (Brobeck et al., 1943; Brooks & Lambert, 1946; Kennedy, 1950). During the dynamic phase, resection of most of the stomach does not appreciably alter the hyperphagia (Brooks, Lockwood, & Wiggins, 1946). Additional lesions of the lateral hypothalamus, however, not only eliminate the hyperphagia but also cause the animal to stop eating altogether (Anand & Brobeck, 1951b).

Once it had been established that animals with ventromedial hypothalamic lesions become obese primarily because they overeat, it became important to specify the cause of the overeating, and this proved to be more difficult. Brobeck (1946) has suggested that these lesions disturb some hypothalamic mechanism which normally maintains a balance between energy need and food intake, but this is not so much an explanation as a description of a generally recognized phenomenon—hyperphagic animals are deficient in the ability to adjust to changed caloric requirements. Thus, although they are less active than normal animals (Hetherington, 1941; Hetherington & Ranson, 1942c), hyperphagic animals that are in the dynamic phase consume about twice as much food as normals (Tepperman et al., 1941). Also, hyperphagic animals that are in the static phase (when food consumption is approximately normal), if fed a diet adulterated with some inert substance such as kaolin or cellulose (Kennedy, 1950; Larsson & Ström, 1957; Teitelbaum, 1955) or subjected to low temperatures (Fregly, Marshall, & Mayer, 1957), do not compensate for the decreased calories or the increased need for heat produc-

tion by increasing their food intake as normals do under these conditions. It should be noted, however, that the failure of obese hypothalamic hyperphagic animals to eat more of a cellulose adulterated diet can also be explained on the grounds that hyperphagic animals are more responsive than normals to alterations in the taste and texture of food; this will be discussed in more detail later.

The overeating which follows ventromedial hypothalamic lesions has also been attributed to a change in emotionality. Wheatley (1944) pointed out that lesions in this area make animals savage; this prompted Brooks, Lockwood, and Wiggins (1946) to suggest that the voraciousness of hypothalamic hyperphagic animals was actually a substitute for attack, but the explanation does not seem to have been generally accepted. It does not account for the overeating which results from lesions in other areas such as the caudal hypothalamus and caudal thalamus where taming effects have been noted (Brobeck, 1946; Hetherington & Ranson, 1942c; Patton, Ruch, & Walker, 1944), and this objection was cited by Brooks (1947) a year later when he repudiated his own earlier position.

Perhaps the most popular explanation is that hyperphagia is a "release phenomenon." Hyperphagic animals overeat not because they are "hungrier" but because, lacking a "satiety" mechanism, they are unable to stop eating once they have started (Anand & Brobeck, 1951a; Kennedy, 1950; Miller, Bailey, & Stevenson, 1950; Teitelbaum, 1955). How this satiety mechanism operates in normal animals is not generally agreed upon. According to Brobeck (1957), the ingestion of food causes an almost immediate increase in heat production which, in turn, causes thermal receptors in the hypothalamus to inhibit eating. Other investigators, however, have suggested that the control of food intake depends

on chemoreceptors situated in the hypothalamus which are sensitive to changes in blood composition (Kennedy, 1953; Mayer, Vitale, & Bates, 1951). Whatever the mechanism involved, the evidence for a satiety center derives from a number of sources, and these will be considered separately.

One line of evidence concerns the difference between the eating patterns of normal and hypothalamic hyperphagic animals. Increased consumption by hyperphagic animals is reflected mainly in meals which are larger than normal (Brooks, Lockwood, & Wiggins, 1946; Larsson & Ström, 1957; Teitelbaum & Campbell, 1958). On a solid or liquid diet, hyperphagic animals, unlike normals, consume as much during the day as at night (Anliker & Mayer, 1956; Brooks, 1946; Teitelbaum & Campbell, 1958), and, on a diet of wet mash, hyperphagic rats eat most just after the food has been changed regardless of the hour (Brobeck et al., 1943; Brooks, Lockwood, & Wiggins, 1946).

A second line of evidence is related to the dissociation between two measures of food drive in hypothalamic hyperphagic animals. These animals do not work as hard as normals on food motivated tasks, even when the deprivation is extended up to 96 hours, but they overeat when fed ad lib. (Ingram, 1958; Miller et al., 1950; Teitelbaum, 1957), and these findings have been interpreted to mean that ventromedial hypothalamic lesions lower food drive but impair the mechanism which inhibits eating. It is interesting to note that the degree of the performance deficit is a function of the amount of work involved. In an experiment in which food was obtained by pushing the lid off a box, hyperphagic animals ate more than normals when the lid was unweighted but less than normals when the lid was weighted (Miller et al., 1950). In another experiment in which

the total daily food supply was obtained by bar pressing during a 12-hour period in a Skinner box, hyperphagic rats pressed more and obtained more food than normals when the ratio of nonreinforced to reinforced responses was low, but they pressed less and obtained less food when the ratio was high (Teitelbaum, 1957).

According to a third line of reasoning, the "release" brought about by ventromedial hypothalamic lesions consists of an overreactivity to stimuli associated with feeding. Thus, stomach preloads (Smith, Salisbury, & Weinberg, 1961) and appetite depressant drugs (Epstein, 1959; Reynolds, 1959) significantly lower food intake in both hyperphagic and normal rats, but the effects of such procedures are more marked in hyperphagic than in normal animals. Hyperphagic animals also overreact to the quality of the diet offered them. Hetherington and Ranson (1942c) noted that hyperphagic animals gained more on a soft diet than on pellets; Miller et al. (1950) observed that a high fat synthetic diet significantly increased the food intake of obese rats that had not appeared to be hyperphagic on a diet of dry chow; and Teitelbaum (1955) reported that the addition of dextrose to a powdered diet significantly increased the food intake of obese hyperphagic animals. On the negative side, the addition of quinine to the diet in an amount which did not disturb normal animals resulted in a significant decrease in food intake by obese hyperphagic rats (Miller et al., 1950; Teitelbaum, 1955). The rejection by obese hyperphagic animals of diets adulterated with cellulose or kaolin can also be explained on this basis, as noted earlier.

One qualifying point should be noted with respect to the extreme sensitivity of hyperphagic animals to the quality of the diet—animals that have not yet be-

come obese are less responsive than obese animals to the taste and texture of food. Kennedy (1950) found that the addition of kaolin to the diet was readily compensated for by recent operates but not by rats that had already become obese; and Teitelbaum (1955) found that, whereas both groups overreacted to cellulose adulteration, the decrease in food intake was more extreme in the obese group. Similarly, the addition of quinine to the diet in an amount which caused the obese group to limit its intake severely had as little effect on the nonobese as on the normal control group (Teitelbaum, 1955). Thus, the hyperreactivity of the hyperphagic animal to the quality of the diet is in some way related to the state of the animal's fat deposits.

### *Aphagia*

We turn now to a group of experiments in which hypothalamic lesions produced a totally different effect on eating. Injury to the lateral hypothalamus at the same rostrocaudal level as the ventromedial nucleus has been found to result in a condition known as "aphagia," which is characterized by refusal to eat and loss of weight. In many cases, death by starvation occurs. The change in eating behavior does not appear to be attributable to a decrease in activity (Anand & Brobeck, 1951a). Aphagia, like hyperphagia, has been demonstrated in a variety of animal species—rats and cats (Anand & Brobeck, 1951a, 1951b; Teitelbaum & Stellar, 1954), monkeys (Anand, Dua, & Shoenberg, 1955), and chickens (Feldman, Larsson, Dimick, & Lepkovsky, 1957). Within the lateral hypothalamus, little damage is required to produce aphagia, and it is sometimes sufficient merely to lower an electrode without turning on the current (Morgane, 1961c; Morrison & Mayer, 1957a, 1957b).

When hypothalamic aphagia was first

described, the failure to eat spontaneously appeared to be permanent and it was thought to result from destruction of an extremely limited area (Anand & Brobeck, 1951a, 1951b). However, recent work has shown that neither of these original observations was entirely accurate. Anand and Brobeck (1951b) found the effective area within the lateral hypothalamus to be so circumscribed that deviations by 1 mm. or less in any direction from their original electrode placement produced only a temporary depression in food intake which did not lead to death. Morgane (1961d) has verified their findings with respect to lesions placed anterior or posterior to the tuberal level, but he finds that, at the tuberal level, the effective region is more extensive, both dorsally and laterally, than was originally supposed. Morgane (1961a) finds, further, that aphagia also results when the lesions are at the same level but are outside the hypothalamus and are restricted to the globus pallidus. Current research has also modified earlier notions concerning the permanence of the effect. Although lateral hypothalamic lesions can lead to death by starvation, recovery is possible, provided the animals are tube fed during the period (which may be several months) when they refuse to eat spontaneously (Morrison & Mayer, 1957a; Teitelbaum & Stellar, 1954). Teitelbaum and his co-workers have studied the recovery pattern in a series of experiments (Epstein & Teitelbaum, 1960; Teitelbaum & Epstein, 1962; Teitelbaum & Stellar, 1954; Williams & Teitelbaum, 1959) and they note that recovery involves four separate stages. In the first, the animal does not eat or drink; in the second, it eats small quantities of preferred foods but does not drink; in the third, it eats normally but still refuses water; and in the fourth, it eats and drinks normally.

### *Dual-Center Theory*

The most widely accepted explanation of the drastic and apparently opposite effects of medial and lateral hypothalamic lesions is the dual-center theory proposed by Anand and Brobeck (1951a, 1951b; see also Anand et al., 1955). According to the theory (which is an expansion of the concept of the "satiety mechanism" previously outlined), the lateral hypothalamus is the site of an "excitatory" or "feeding" center which initiates the feeding response when the animal is stimulated by the sight or smell of food, and the ventromedial nucleus is the site of a corresponding "inhibitory" or "satiety" center, which stops the response when the animal has satisfied its physiological need for food. Although the theory cannot account satisfactorily for all the data, it does account for a great deal.

Perhaps the strongest support for the dual-center theory comes from the finding that ablation and electrical stimulation of the ventromedial nucleus and of the lateral hypothalamic area have reciprocal effects. Thus, removal of the inhibitory center results in an increase in food consumption (Hetherington & Ranson, 1940), but stimulation of the same area results in a decrease in food consumption (Olds, 1958; Smith, 1956) and in the inhibition of a previously learned instrumental response motivated by food reward (Wyrwicka & Dobrzecka, 1960). Similarly, ablation of the excitatory center causes cessation of eating (Anand & Brobeck, 1951a, 1951b), but its stimulation causes increased food intake (Delgado & Anand, 1953; Larsson, 1954; Olds, 1958; Smith, 1956). Alterations in food intake, along the lines predicted by the theory, have also been produced by injection of drugs directly into the hypothalamus. This method is of particular interest since it produces more specific effects than electrical stimula-

tion; it is possible, for example, by injecting the appropriate drug at the same hypothalamic site, to alter not only the type of response (i.e., eating versus drinking; Grossman, 1960) but also the direction of the response (i.e., increase versus decrease in food intake; Epstein, 1960).

Additional support for the theory comes from the finding that the eating elicited by stimulation of the lateral hypothalamus is motivated and not merely reflexive. Satiated animals, whether stimulated electrically or chemically, will perform a variety of instrumental responses to get food. Coons and Miller (cited by Miller, 1957) trained rats to push back a hinged door; Grossman (cited by Miller, 1960) trained rats to bar press; Morgane (1961b) trained rats to cross an electrified grid; and Wyrwicka, Dobrzecka, and Tarnecki (1959) trained goats to place the foreleg on a food tray. Wyrwicka and Dobrzecka (1960) further showed that, once this conditioned response had been elicited in satiated goats by stimulation of the lateral hypothalamus, it could then be inhibited by subsequent stimulation of the ventromedial nucleus.

Both Anand (Sharma, Anand, Dua, & Singh, 1961) and Brobeck (Brobeck, Larsson, & Reyes, 1956) have been associated with attempts to demonstrate the existence of a dual hypothalamic feeding mechanism in yet another way, by recording electrical activity in the medial and lateral hypothalamus under varying conditions of hunger and satiety, but these studies have not furnished unambiguous support for the theory. On the basis of their results, Brobeck and his co-workers concluded that amphetamine, a known appetite depressant, acts upon the medial hypothalamus. However, subsequent attempts by both Epstein (1959) and Reynolds (1959) to demonstrate this behaviorally failed.

Both investigators found that the drug suppressed eating in hypothalamic hyperphagic rats as well as in normals. The evidence furnished by Sharma et al. (1961) appears to be more convincing, but it should be noted that, in the latter experiment, as well as in the one by Brobeck and his group, the data consist of limited selections from raw electroencephalogram records; in the absence of quantification, what constitutes a "significant change" in electrical activity may be a matter of dispute.

Perhaps the major objection to the dual-center theory is that it does not entirely account for the complexity of the changes that follow ventromedial and lateral hypothalamic lesions. Consider, for example, the problem of what happens when an animal "recovers" from hyperphagia. The lateral hypothalamus is believed to be the more important of the two centers, and it is usually assumed that, when the hyperphagic animal enters the static phase, inhibition of feeding is mediated by undamaged tissue in this area (Brobeck, 1957). That the new control over food intake is not simply a matter of vicarious function following a more or less fixed recovery period is shown by a number of facts—if animals are prevented from overeating for a period of several months following operation, they become obese when placed on an unrestricted feeding schedule (Brobeck et al., 1943); hyperphagia can be reinstated in animals that have presumably recovered by dieting the animals down to a normal weight level (Brobeck et al., 1943); and, when ad lib. feeding is resumed, the hyperphagic tapers off at a higher weight level than previously (Brooks & Lambert, 1946). That body weight somehow plays a role in the regulation of food intake in hyperphagics has already been noted, but how it interacts with a central inhibitory mechanism is not clear. It should also

be noted that Anand and Brobeck's (1951a, 1951b) reason for assuming the lateral center to be more basic was that aphagia resulted when both medial and lateral destruction occurred in the same animal, and this was true whether the medial area was destroyed first or the two areas were destroyed simultaneously. Yet, it has recently been shown that, when such animals recover from aphagia, they become obese (Teitelbaum & Epstein, 1962). What determines this sequence of events is, as yet, unexplained. Another point to be considered is that the aphagia that occurs following lateral hypothalamic lesions is not an all-or-none phenomenon, as one would expect if the lesions simply removed an excitatory center. Instead, whether recently operated animals starve to death or maintain their normal weight levels depends upon whether they are offered dry food or a liquid diet (Williams & Teitelbaum, 1959). Even after they have presumably recovered, such animals will sometimes refuse quinine adulterated foods at dosage levels accepted by normal animals (Teitelbaum & Epstein, 1962); the fact that the same behavior has been noted in recovered hyperphagics (Teitelbaum, 1955) suggests that there is some underlying relationship between hyperphagia and aphagia. It is conceivable that they represent different degrees or different stages of the same disorder—an extreme sensitivity to the taste and texture of food.

Morgane (1961a, 1961b, 1961c, 1961d) has criticized Anand and Brobeck's dual-center theory on other grounds. He suggests that the lateral hypothalamus contains not one feeding center but two—a midlateral "motivational" system, dependent upon the integrity of the medial forebrain bundle, and a more basic far-lateral "metabolic" system, dependent upon the integrity of pallidofugal fibers. The distinction is

based, in part, on Morgane's supposed demonstration of a nonreversible aphagia following far-lateral but not following midlateral lesions. Actually, as Teitelbaum and Epstein (1962) have shown, all animals with lateral hypothalamic lesions recover eventually. However, the recovery period is longer after far-lateral lesions, and some animals with such lesions never drink water spontaneously. Morgane's findings were apparently due to his failure to distinguish between two effects—loss of spontaneous eating and failure to eat because of lack of adequate hydration. The other basis on which Morgane claims to distinguish between motivational and metabolic centers is his own demonstration that far-lateral stimulation elicits both feeding and food-motivated behavior only as long as the medial forebrain bundle is intact. With the medial forebrain bundle sectioned, such stimulation elicits feeding but not food-motivated behavior. In these experiments, the area stimulated was extremely close to the damaged tissue and the effect could have been due to this factor alone, a possibility for which Morgane has not furnished suitable controls. Despite the fact that there does not seem to be sufficient evidence for two feeding centers in the lateral hypothalamus, Morgane's data are of considerable interest to psychologists, especially since he has furnished a more careful anatomical analysis of the lateral hypothalamic syndrome than has been attempted to date.

A final and less serious objection to the dual-center theory is that the only parts of the hypothalamus that are supposed to function in the regulation of food intake are the ventromedial and lateral parts at the tuberal level. It will be noted that the present discussion has been confined, so far, to experiments dealing with injuries or stimulation at this level, but there are a few experi-

ments implicating other parts of the hypothalamus in feeding behavior, and these will now be described briefly.

Lesions placed rostral to the ventromedial hypothalamic nucleus produce obesity in dogs (Biggart & Alexander, 1939; Heinbecker & White, 1942; Heinbecker, White, & Rolf, 1944), although they have no effect on food intake in rats, cats, or monkeys (Anand & Brobeck, 1951a; Anand et al., 1955; Hetherington, 1944). With respect to lesions placed in the caudal hypothalamus, the evidence is confusing. There are several studies in which lesions in the region of the mammillary bodies produced aphagia (Clark, Magoun, & Ranson, 1939; Ingram, Barris, & Ranson, 1936; Ranson, 1939). On the other hand, Hetherington and Ranson (1942b) produced obesity in rats by ablating a part of the lateral hypothalamus dorsolateral to the mammillary bodies, and they attributed their results to the destruction of fibers from the ventromedial nucleus which pass through this region on the way to the midbrain. In other studies, caudal hypothalamic lesions in rats and cats (Anand & Brobeck, 1951a) and in cats and monkeys (Anand et al., 1955) produced no change in food intake. To confuse the issue further, in one study in which caudal hypothalamic lesions eliminated spontaneous eating, it was reported that the animals chewed and swallowed voraciously if food was forced into their mouths (Ingram et al., 1936).

Recent studies utilizing the electrical stimulation technique have also implicated parts of the hypothalamus not previously believed to be involved in the regulation of food intake, and here the evidence is more consistent. Eating in satiated animals was induced by stimulating the preoptic region (Robinson & Mishkin, 1962) and the caudal hypothalamus in the vicinity of the mammillary bodies (Larsson, 1954; Maire,

1956; Ruch, Maire, & Patton, 1956). Also, cats stimulated in the dorsal hypothalamic nucleus or in the paraventricular hypothalamic nucleus performed a learned response to obtain food (Grastyán, Lissák, & Kékesi, 1956). Some of these data are inconclusive because no histological verification of placements has been presented. It is also possible that some of the results can be "explained away" on the basis of current spread to the feeding or satiety centers. Nevertheless, until the issue is resolved by further research, the possibility that other parts of the hypothalamus function in the regulation of food intake cannot be entirely ignored.

#### EFFECTS OF EXTRAHYPOTHALAMIC LESIONS ON HUNGER

In this section, only those instances of increased or decreased food intake following brain lesions will be reported that bear some resemblance to hypothalamic hyperphagia or aphagia. Omitted are reports of increased food intake which occur in conjunction with increased energy requirements, as in the case of frontal lobe lesions (e.g., Richter & Hawkes, 1939), as well as reports of decreased food intake in which the behavioral descriptions suggest that motor or intellectual deficits were responsible for the feeding change (e.g., Bailey & Davis, 1942; Kennard, 1955). It should be noted that a great deal of the material which follows was derived from experiments in which feeding changes were reported only incidentally; and, consequently, in many cases, systematic records of food intake and of weight gain were not reported. This admittedly limits the conclusions that can be drawn about the effects of lesions in any one area but, taken as a group, these experiments are important because they call attention to a topic which has not yet been adequately explored, the part

played by areas other than the hypothalamus in the regulation of food intake.

Destruction of the dorsomedial thalamic nucleus reportedly produces a syndrome which is virtually identical to the one that follows ventromedial hypothalamic injury. Cats subjected to such an operation were found to be hyperphagic, irritable, and underactive. Also, as was the case with ventromedial hypothalamic animals, there seemed to be a relation between eating and emotionality. When the cats were subjected to a mild stress, such as having their tails pinched, they gulped food.

It appeared as though the animals released their savage attacks upon the meat rather than turning to attack the offending stimulus as would be done, in most instances, by unoperated cats [Schreiner, Rioch, Pechtel, & Masserman, 1953, p. 239].

Posterior thalamic lesions extending into the rostral mesencephalon also increase eating, although, in this case, the feeding change is accompanied by tameness rather than by savageness. In the course of a series of experiments aimed at establishing the arcuate nucleus as the thalamic relay for taste, Ruch and his co-workers noted incidentally that a few operate monkeys that showed only a mild and transient taste deficit tripled their caloric intake and ultimately became obese (Patton et al., 1944; Ruch, Blum, & Brobeck, 1941; Ruch, Patton, & Brobeck, 1942). Lesions confined to the rostral mesencephalon may also produce obesity provided they are sufficiently extensive to interrupt most of the lateral sensory tracts (Sprague, Chambers, & Stellar, 1961). Small lesions in the same area have a somewhat different effect—*ad lib.* food intake remains unchanged but motivation to work for food is substantially decreased (Ehrlich, 1960).

The notion that temporal lobe struc-

tures might be involved in the regulation of food intake has received a fair amount of attention in recent years. The idea is an old one, however, and dates back at least as far as 1888, when Brown and Schäfer noted that one monkey with a complete bitemporal lobectomy was a hyperphagic and indiscriminate feeder. Over a half century later, Terzian and Ore (1955) observed the same effect in the case of a human subjected to a similar operation:

Some days after the operation the patient demonstrated an insatiable appetite and ate at least as much as four normal persons. He prepared himself for meals as for a ceremony. He would personally go to the kitchen and insistently ask for food at any hour. He would look for a secluded corner far from anyone, eat everything voraciously without preference for any certain food, lick the dish incessantly, and after 15 minutes asked [sic] for more food [p. 375].

Hyperphagia results also from more restricted damage to the temporal lobe. This was first shown by Brown and Schäfer (1888), who removed only the superior temporal gyri in one monkey. Recently, the effect has also been produced by ablation of the entorhinal area. Phalangers, which are normally nocturnal feeders, were found to eat ravenously at any time of day after operation; the change was only temporary, however, and disappeared entirely within about 2 weeks (Adey, Merrillees, & Sunderland, 1956).

The temporal lobe structure most often implicated in feeding changes is the amygdala and, with some exceptions (Anand & Brobeck, 1952; Kling & Schwartz, 1961; Schreiner & Kling, 1953), the effect commonly produced is hyperphagia. This change has been observed not only following operations in which the amygdala was damaged along with other temporal lobe structures but also following operations restricted to the amygdala, and it has been reported

in a variety of species—monkeys (Schwartzbaum, 1961), cats (Green, Clemente, & de Groot, 1957; Morgane & Kosman, 1957, 1959; Wood, 1958), a baboon (Adey, 1958), and dogs (Fuller, Rosvold, & Pribram, 1957). In the latter case, however, the change was temporary. The rat is the only common laboratory animal in which hyperphagia has not been reported following amygdalectomy; instead, the operation seems to produce an aphagia of variable duration (Anand & Brobeck, 1952; Kling & Schwartz, 1961).

Presently available evidence suggests that amygdaloid hyperphagia is not as striking a phenomenon as hypothalamic hyperphagia. In the only studies to date in which long-term records were kept of preoperative and postoperative food intake and weight gain, Morgane and Kosman (1957, 1959) found that, while amygdalectomized cats ate significantly more food and gained significantly more weight than normal controls, the increases in weight did not go beyond 50% of the preoperative levels; hypothalamic hyperphagics, on the other hand, frequently double their initial weight (Tepperman et al., 1941). The dissociation between consummatory and motivated behavior that occurs in animals made hyperphagic by ventromedial hypothalamic lesions has been shown to occur also in animals made hyperphagic by amygdaloid lesions. Schwartzbaum (1960, 1961) has reported that amygdalectomized monkeys that were hyperphagic on an ad lib. diet were relatively insensitive to increasing amounts of food deprivation and to shifts in the size of the food reward in a situation in which they were required to bar press for food.

Further evidence concerning the role played by extrahypothalamic structures in the control of feeding comes from some recent studies utilizing electrical

stimulation. It is interesting to note that eating was elicited by neocortical stimulation in only one instance (Delgado, 1952); all the other cases in which feeding behavior was altered involved stimulation of structures belonging to, or associated with, the limbic system. Thus, eating in satiated animals has been elicited by stimulation of the olfactory area, anterior thalamus, septum, cingulate gyrus, hippocampal commissure, amygdala, and midbrain (Grimm, 1960; Miller, 1960; Robinson & Mishkin, 1962; Smith, McFarland, & Teitelbaum, 1961; Wyrwicka & Doty, 1962). Stimulation of different parts of some of these structures has also been observed to inhibit food intake (Fonberg & Delgado, 1961; Robinson & Mishkin, 1962).

#### FEEDING SYSTEM

When Bailey and Bremer (1921) described the effects of hypothalamic lesions on food intake, little was known about this structure, and they were careful to specify that it was a "region known to anatomists as the hypothalamus [p. 774]." They enthusiastically predicted that "the time is not far distant when the neuropathologist will no more think of omitting to examine the hypothalamus than he would the motor cortex [p. 802]." Their prediction was accurate, and currently the pendulum has swung so far the other way that the hypothalamus has been credited with containing excitatory and inhibitory centers not only for feeding but for each type of motivation (Stellar, 1954).

From the studies reviewed in the preceding section, it is apparent that changes in food intake occur as a result of ablating or stimulating many parts of the brain, and it would seem, therefore, that the hypothalamus is part of a larger "feeding system" that also includes parts of the thalamus, temporal lobe,

frontal lobe, and midbrain. (Of course, it is also possible that more than one system is involved, with each one functioning more or less independently. However, since current research has not progressed far enough to make this a serious possibility, the more parsimonious explanation, in terms of a single system, is preferable.) Whether one takes the further step of viewing the hypothalamus as a unique part of this system is another matter. In their initial formulation of the dual-center theory, Anand and Brobeck paid scant attention to the possibility that structures other than the hypothalamus might be involved in the regulation of food intake, but in a recent article, Anand (1959) notes what he terms "cerebral influences" on hunger. However, he suggests that these are secondary and are mediated by way of the hypothalamus, which is still considered to be the site of primary hunger receptors. A more moderate view (and one that seems to accord better with the known facts) is that the importance of the hypothalamus lies in the fact that it is a point of convergence for a great many fiber systems, all of which act on lower reflex arcs controlling feeding (Morgane, 1961a). One would, therefore, expect the effects of hypothalamic lesions to be more severe simply because there would be fewer alternate pathways available following hypothalamic than following other types of brain injury.

Neural control of food intake remains a challenging area for investigation. Research to date has progressed only to the stage of mapping crucial areas of the brain, but a great deal remains to be learned especially with regard to such problems as the following: (a) The relationship between hyperphagia and aphagia. In the case of hypothalamic lesions, it was noted that both syndromes involved alterations in taste but whether

this is also true when food intake is altered by extrahypothalamic lesions is not known. (b) The relationship between food and water regulation. The extensive (and often contradictory) literature on the subject of neural control of water intake will not be reviewed here, but it should be noted that, in many cases, alterations in food intake or in food motivation are accompanied by similar changes in water intake or in water motivation (Delgado, 1952; Ehrlich, 1960; Morrison & Mayer, 1957a, 1957b; Paget, 1897; Robinson & Mishkin, 1962; Ruch et al., 1956; Schreiner & Kling, 1953; Teitelbaum & Epstein, 1962; Teitelbaum & Stellar, 1954; Williams & Teitelbaum, 1959). (c) The discrepancy between consummatory and motivated behavior that sometimes occurs as a result of brain injury. The phenomenon has been noted following hypothalamic (Miller et al., 1950; Teitelbaum, 1957), amygdaloid (Schwartzbaum, 1960, 1961), and midbrain (Ehrlich, 1960) lesions. In all of the foregoing experiments, the possibility that the motivational change was not related to feeding as such but to some more general change was not adequately explored. Yet, some preliminary data reported by Ingram (1958), showing that avoidance as well as food motivation is impaired in hypothalamic hyperphagic cats, makes this a reasonable possibility. (d) The relation between hypothalamic and extrahypothalamic structures involved in the regulation of food intake. Multiple lesion studies or combined lesion and stimulation techniques should help to answer the question of whether the various alternate pathways function independently.

#### REFERENCES

- ADEY, W. R. Organization of the rhinencephalon. In H. H. Jasper et al. (Eds.), *Reticular formation of the brain*. Boston: Little, Brown, 1958. Pp. 621-644.

- ADEY, W. R., MERRILLEES, N. C. R., & SUNDERLAND, S. The entorhinal area: Behavioural, evoked potential, and histological studies of its interrelationships with brain-stem regions. *Brain*, 1956, 79, 414-439.
- ANAND, B. K. Higher nervous control over food intake. In, *XXI International Physiological Congress*, 1959. Pp. 196-202.
- ANAND, B. K., & BROBECK, J. R. Hypothalamic control of food intake in rats and cats. *Yale J. Biol. Med.*, 1951, 24, 123-140. (a)
- ANAND, B. K., & BROBECK, J. R. Localization of a "feeding center" in the hypothalamus of the rat. *Proc. Soc. Exp. Biol. Med.*, N. Y., 1951, 77, 323-324. (b)
- ANAND, B. K., & BROBECK, J. R. Food intake and spontaneous activity of rats with lesions in the amygdaloid nuclei. *J. Neurophysiol.*, 1952, 15, 421-430.
- ANAND, B. K., DUA, S., & SHOENBERG, KATE. Hypothalamic control of food intake in cats and monkeys. *J. Physiol.*, 1955, 127, 143-152.
- ANLIKER, J., & MAYER, J. An operant conditioning technique for studying feeding-fasting patterns in normal and obese mice. *J. appl. Physiol.*, 1956, 8, 667-670.
- BAILEY, P., & BREMER, F. Experimental diabetes insipidus. *Arch. intern. Med.*, 1921, 28, 773-803.
- BAILEY, P., & DAVIS, E. W. Effects of lesions of the periaqueductal gray matter in the cat. *Proc. Soc. Exp. Biol. Med.*, N. Y., 1942, 51, 305-306.
- BIGGART, J. H., & ALEXANDER, G. L. Experimental diabetes insipidus. *J. Pathol. Bacteriol.*, London, 1939, 48, 405-425.
- BROBECK, J. R. Mechanism of the development of obesity in animals with hypothalamic lesions. *Physiol. Rev.*, 1946, 26, 541-559.
- BROBECK, J. R. Neural control of hunger, appetite, and satiety. *Yale J. Biol. Med.*, 1957, 29, 565-574.
- BROBECK, J. R., LARSSON, S., & REYES, E. A study of the electrical activity of the hypothalamic feeding mechanism. *J. Physiol.*, 1956, 132, 358-364.
- BROBECK, J. R., TEPPERMAN, J., & LONG, C. N. H. Experimental hypothalamic hyperphagia in the albino rat. *Yale J. Biol. Med.*, 1943, 15, 831-853.
- BROOKS, C. McC. The relative importance of changes in activity in the development of experimentally produced obesity in the rat. *Amer. J. Physiol.*, 1946, 147, 708-716.
- BROOKS, C. McC. Appetite and obesity. *N. Z. med. J.*, 1947, 46, 243-254.
- BROOKS, C. McC., & LAMBERT, E. F. A study of the effect of limitation of food intake and the method of feeding on the rate of weight gain during hypothalamic obesity in the albino rat. *Amer. J. Physiol.*, 1946, 147, 695-707.
- BROOKS, C. McC., LAMBERT, E. F., & BARD, P. Experimental production of obesity in the monkey (*Macaca mulatta*). *Federat. Proc.*, 1942, 1, 11. (Abstract)
- BROOKS, C. McC., LOCKWOOD, R. A., & WIGGINS, M. L. A study of the effect of hypothalamic lesions on the eating habits of the albino rat. *Amer. J. Physiol.*, 1946, 147, 735-741.
- BROOKS, C. McC., MARINE, D. N., & LAMBERT, E. F. A study of the food-feces ratio and of the oxygen consumption of albino rats during various phases of experimentally produced obesity. *Amer. J. Physiol.*, 1946, 147, 717-726.
- BROWN, S., & SCHÄFER, E. A. An investigation into the functions of the occipital and temporal lobes of the monkey's brain. *Phil. Trans. Roy. Soc. London, Ser. B*, 1888, 179, 303-327.
- CLARK, G., MAGOUN, H. W., & RANSON, S. W. Hypothalamic regulation of body temperature. *J. Neurophysiol.*, 1939, 2, 61-80.
- DELGADO, J. M. R. Responses evoked in waking cat by electrical stimulation of motor cortex. *Amer. J. Physiol.*, 1952, 171, 436-446.
- DELGADO, J. M. R., & ANAND, B. K. Increase of food intake induced by electrical stimulation of the lateral hypothalamus. *Amer. J. Physiol.*, 1953, 172, 162-168.
- EHRLICH, ANNETTE. Effects of tegmental lesions on motivated behavior in rats. Unpublished doctoral dissertation, McGill University, 1960.
- EPSTEIN, A. N. Suppression of eating and drinking by amphetamine and other drugs in normal and hyperphagic rats. *J. comp. physiol. Psychol.*, 1959, 52, 37-45.
- EPSTEIN, A. N. Reciprocal changes in feeding behavior produced by intrahypothalamic chemical injections. *Amer. J. Physiol.*, 1960, 199, 969-974.
- EPSTEIN, A. N., & TEITELBAUM, P. Role of adipisia in lateral hypothalamic starvation. *Science*, 1960, 132, 1491. (Abstract)
- FELDMAN, S. E., LARSSON, S., DIMICK, M. K., & LEPKOVSKY, S. Aphagia in chickens. *Amer. J. Physiol.*, 1957, 191, 259-261.
- FONBERG, ELZBIETA, & DELGADO, J. M. R. Inhibitory effects of amygdala on food intake and conditioning. *Federat. Proc.*, 1961, 20, 335. (Abstract)
- FREGLY, M. J., MARSHALL, N. B., & MAYER, J.

- Effect of changes in ambient temperature on spontaneous activity, food intake and body weight of goldthioglucose-obese and non-obese mice. *Amer. J. Physiol.*, 1957, 188, 435-438.
- FULLER, J. L., ROSVOLD, H. E., & PRIBRAM, K. H. The effect on affective and cognitive behavior in the dog of lesions of the pyriform-amygdala-hippocampal complex. *J. comp. physiol. Psychol.*, 1957, 50, 89-96.
- GRASTYÁN, E., LISSÁK, K., & KÉKESI, F. Facilitation and inhibition of conditioned alimentary and defensive reflexes by stimulation of the hypothalamus and reticular formation. *Acta physiol. Hung.*, 1956, 9, 133-151.
- GREEN, J. D., CLEMENTE, C. D., & DE GROOT, J. Rhinencephalic lesions and behavior in cats. *J. comp. Neurol.*, 1957, 108, 505-545.
- GRIMM, R. J. Feeding behavior and electrical stimulation of the brain of *Carassius auratus*. *Science*, 1960, 131, 162-163.
- GROSSMAN, S. P. Eating or drinking elicited by direct adrenergic or cholinergic stimulation of hypothalamus. *Science*, 1960, 132, 301-302.
- HEINBECKER, P., & WHITE, H. L. Experimental obesity in the dog. *Proc. Soc. Exp. Biol. Med.*, N. Y., 1942, 49, 324-327.
- HEINBECKER, P., WHITE, H. L., & ROLF, DORIS. Experimental obesity in the dog. *Amer. J. Physiol.*, 1944, 141, 549-565.
- HETHERINGTON, A. W. The relation of various hypothalamic lesions to adiposity and other phenomena in the rat. *Amer. J. Physiol.*, 1941, 133, 326-327. (Abstract)
- HETHERINGTON, A. W. The production of hypothalamic obesity in rats already displaying chronic hypopituitarism. *Amer. J. Physiol.*, 1943, 140, 89-92.
- HETHERINGTON, A. W. Non-production of hypothalamic obesity in the rat by lesions rostral or dorsal to the ventro-medial hypothalamic nuclei. *J. comp. Neurol.*, 1944, 80, 33-45.
- HETHERINGTON, A. W., & RANSON, S. W. Experimental hypothalamico-hypophyseal obesity in the rat. *Proc. Soc. Exp. Biol. Med.*, N. Y., 1939, 41, 465-466.
- HETHERINGTON, A. W., & RANSON, S. W. Hypothalamic lesions and adiposity in the rat. *Anat. Rec.*, 1940, 78, 149-172.
- HETHERINGTON, A. W., & RANSON, S. W. Effect of early hypophysectomy on hypothalamic obesity. *Endocrinology*, 1942, 31, 30-34. (a)
- HETHERINGTON, A. W., & RANSON, S. W. The relation of various hypothalamic lesions to adiposity in the rat. *J. comp. Neurol.*, 1942, 76, 475-499. (b)
- HETHERINGTON, A. W., & RANSON, S. W. The spontaneous activity and food intake of rats with hypothalamic lesions. *Amer. J. Physiol.*, 1942, 136, 609-617. (c)
- INGRAM, W. R. Modification of learning by lesions and stimulation in the diencephalon and related structures. In H. H. Jasper et al. (Eds.), *Reticular formation of the brain*. Boston: Little, Brown, 1958. Pp. 535-544.
- INGRAM, W. R., BARRIS, R. W., & RANSON, S. W. Catalepsy: An experimental study. *Arch. Neurol. Psychiat.*, Chicago, 1936, 35, 1175-1197.
- KENNARD, MARGARET A. Effect of bilateral ablation of cingulate area on behaviour of cats. *J. Neurophysiol.*, 1955, 18, 159-169.
- KENNEDY, G. C. The hypothalamic control of food intake in rats. *Proc. Roy. Soc., London, Ser. B*, 1950, 137, 535-549.
- KENNEDY, G. C. The role of depot fat in the hypothalamic control of food intake in the rat. *Proc. Roy. Soc., London, Ser. B*, 1953, 140, 578-592.
- KLING, A., & SCHWARTZ, NEENA B. Effects of amygdectomy on feeding in infant and adult animals. *Federat. Proc.*, 1961, 20, 335. (Abstract)
- LARSSON, S. On the hypothalamic organisation of the nervous mechanism regulating food intake. *Acta physiol. Scand., Stockholm*, 1954, 32(Suppl. No. 115, Pt. 1), 14-40.
- LARSSON, S., & STRÖM, L. Some characteristics of goldthioglucose obesity in the mouse. *Acta physiol. Scand., Stockholm*, 1957, 38, 298-308.
- MAIRE, F. W. Eating and drinking responses elicited by diencephalic stimulation in unanesthetized rats. *Federat. Proc.*, 1956, 15, 124. (Abstract)
- MARSHALL, N. B., BARNETT, R. J., & MAYER, J. Hypothalamic lesions in goldthioglucose injected mice. *Proc. Soc. Exp. Biol. Med.*, N. Y., 1955, 90, 240-244.
- MAYER, J., FRENCH, ROSALIE G., ZIGHERA, CLAUDINE F., & BARNETT, R. J. Hypothalamic obesity in the mouse: Production, description and metabolic characteristics. *Amer. J. Physiol.*, 1955, 182, 75-82.
- MAYER, J., VITALE, J. J., & BATES, MARGARET W. Mechanism of the regulation of food intake. *Nature, London*, 1951, 167, 562-563.
- MILLER, N. E. Experiments on motivation: Studies combining psychological, physiological, and pharmacological techniques. *Science*, 1957, 126, 1271-1278.
- MILLER, N. E. Motivational effects of brain

- stimulation and drugs. *Federat. Proc.*, 1960, 19, 846-854.
- MILLER, N. E., BAILEY, C. J., & STEVENSON, J. A. F. Decreased "hunger" but increased food intake resulting from hypothalamic lesions. *Science*, 1950, 112, 256-259.
- MOHR. Mittheilungen für neuropathologische Studien: 4. Hypertrophie (Markschwammige Entartung?) der *Hypophysis cerebri* und dadurch bedingter Druck auf die Hirngrundfläche, insbesondere auf die Sehnerven das *Chiasma* derselben und den linkseitigen Hirnschenkel. *Wsch. ges. Heilk.*, 1840, No. 35, 565-571. Cited by J. R. Brobeck, Obesity in animals with hypothalamic lesions. *Physiol. Rev.*, 1946, 26, 541-559.
- MORGANE, P. J. Alterations in feeding and drinking behavior of rats with lesions in *globi pallidi*. *Amer. J. Physiol.*, 1961, 201, 420-428. (a)
- MORGANE, P. J. Distinct "feeding" and "hunger motivating" systems in the lateral hypothalamus of the rat. *Science*, 1961, 133, 887-888. (b)
- MORGANE, P. J. Evidence of a "hunger motivational" system in the lateral hypothalamus of the rat. *Nature, London*, 1961, 191, 672-674. (c)
- MORGANE, P. J. Medial forebrain bundle and "feeding centers" of the hypothalamus. *J. comp. Neurol.*, 1961, 117, 1-25. (d)
- MORGANE, P. J., & KOSMAN, A. J. Alterations in feline behaviour following bilateral amygdalotomy. *Nature, London*, 1957, 180, 598-600.
- MORGANE, P. J., & KOSMAN, A. J. A rhinencephalic feeding center in the cat. *Amer. J. Physiol.*, 1959, 197, 158-162.
- MORRISON, S. D., & MAYER, J. Adipsia and aphagia in rats after lateral subthalamic lesions. *Amer. J. Physiol.*, 1957, 191, 248-254. (a)
- MORRISON, S. D., & MAYER, J. Effect of sham operations in the hypothalamus on food and water intake of the rat. *Amer. J. Physiol.*, 1957, 191, 255-258. (b)
- OLDS, J. Self-stimulation of the brain: Its use to study local effects of hunger, sex, and drugs. *Science*, 1958, 127, 315-324.
- PAGET, S. On cases of voracious hunger and thirst from injury or disease of the brain. *Trans. Clin. Soc. London*, 1897, 30, 113-119.
- PATTON, H. D., RUCH, T. C., & WALKER, A. E. Experimental hypogeusia from Horsley-Clarke lesions of the thalamus in *Macaca mulatta*. *J. Neurophysiol.*, 1944, 7, 171-184.
- RANSON, S. W. Somnolence caused by hypothalamic lesions in the monkey. *Arch. Neurol. Psychiat., Chicago*, 1939, 41, 1-23.
- RANSON, S. W., FISHER, C., & INGRAM, W. R. Adiposity and diabetes mellitus in a monkey with hypothalamic lesions. *Endocrinology*, 1938, 23, 175-181.
- REYNOLDS, R. W. The effect of amphetamine on food intake in normal and hypothalamic hyperphagic rats. *J. comp. physiol. Psychol.*, 1959, 52, 682-684.
- RICHTER, C. P., & HAWKES, C. D. Increased spontaneous activity and food intake produced in rats by removal of the frontal poles of the brain. *J. Neurol. Psychiat.*, 1939, 2, 231-242.
- ROBINSON, B. W., & MISHKIN, M. Alimentary responses evoked from forebrain structures in *Macaca mulatta*. *Science*, 1962, 136, 260-262.
- RUCH, T. C., BLUM, M., & BROBECK, J. Taste disturbances from thalamic lesions in monkeys. *Amer. J. Physiol.*, 1941, 133, 433-434. (Abstract)
- RUCH, T. C., MAIRE, F. W., & PATTON, H. D. Stimulus bound eating and drinking responses elicited by diencephalic stimulation. In *XX International Physiological Congress*, 1956. Vol. 1. P. 788. (Abstract)
- RUCH, T. C., PATTON, H. D., & BROBECK, J. R. Hyperphagia and adiposity in relation to disturbances of taste. *Federat. Proc.*, 1942, 1, 76. (Abstract)
- SCHREINER, L., & KLING, A. Behavioral changes following rhinencephalic injury in cat. *J. Neurophysiol.*, 1953, 16, 643-659.
- SCHREINER, L., RIOCH, McK., PECHTEL, C., & MASSERMAN, J. H. Behavioral changes following thalamic injury in cat. *J. Neurophysiol.*, 1953, 16, 234-246.
- SCHWARTZBAUM, J. S. Changes in reinforcing properties of stimuli following ablation of the amygdaloid complex in monkeys. *J. comp. physiol. Psychol.*, 1960, 53, 388-395.
- SCHWARTZBAUM, J. S. Some characteristics of "amygdaloid hyperphagia" in monkeys. *Amer. J. Psychol.*, 1961, 74, 252-259.
- SHARMA, K. N., ANAND, B. K., DUA, S., & SINGH, B. Role of stomach in regulation of activities of hypothalamic feeding centers. *Amer. J. Physiol.*, 1961, 201, 593-598.
- SMITH, M. H., SALISBURY, R., & WEINBERG, H. The reaction of hypothalamic-hyperphagic rats to stomach preloads. *J. comp. physiol. Psychol.*, 1961, 54, 660-664.
- SMITH, O. A. Stimulation of lateral and medial hypothalamus and food intake in the rat. *Anat. Rec.*, 1956, 124, 363-364. (Abstract)
- SMITH, O. A., MCFARLAND, W. L., & TEITELBAUM, H. Motivational concomitants of eating elicited by stimulation of the an-

- terior thalamus. *J. comp. physiol. Psychol.*, 1961, 54, 484-488.
- SPRAGUE, J. M., CHAMBERS, W. W., & STELLAR, E. Attentive, affective, and adaptive behavior in the cat. *Science*, 1961, 133, 165-173.
- STELLAR, E. The physiology of motivation. *Psychol. Rev.*, 1954, 61, 5-22.
- TEITELBAUM, P. Sensory control of hypothalamic hyperphagia. *J. comp. physiol. Psychol.*, 1955, 48, 156-163.
- TEITELBAUM, P. Random and food-directed activity in hyperphagic and normal rats. *J. comp. physiol. Psychol.*, 1957, 50, 486-490.
- TEITELBAUM, P., & CAMPBELL, B. A. Ingestion patterns in hyperphagic and normal rats. *J. comp. physiol. Psychol.*, 1958, 51, 135-141.
- TEITELBAUM, P., & EPSTEIN, A. N. The lateral hypothalamic syndrome: Recovery of feeding and drinking after lateral hypothalamic lesions. *Psychol. Rev.*, 1962, 69, 74-90.
- TEITELBAUM, P., & STELLAR, E. Recovery from failure to eat produced by hypothalamic lesions. *Science*, 1954, 120, 894-895.
- TEPPERMAN, J., BROBECK, J. R., & LONG, C. N. H. A study of experimental hypothalamic obesity in the rat. *Amer. J. Physiol.*, 1941, 133, 468-469. (Abstract)
- TERZIAN, H., & ORE, G. D. Syndrome of Klüver and Bucy: Reproduced in man by bilateral removal of the temporal lobes. *Neurology*, 1955, 5, 373-380.
- WHEATLEY, M. D. The hypothalamus and affective behavior in cats: A study of the effects of experimental lesions, with anatomic correlations. *Arch. Neurol. Psychiat.*, Chicago, 1944, 52, 296-316.
- WILLIAMS, D. R., & TEITELBAUM, P. Some observations on the starvation resulting from lateral hypothalamic lesions. *J. comp. physiol. Psychol.*, 1959, 52, 458-465.
- WOOD, C. D. Behavioral changes following discrete lesions of temporal lobe structures. *Neurology*, 1958, 8, 215-220.
- WYRWICKA, WANDA, & DOBRZECKA, C. Relationship between feeding and satiation centers of the hypothalamus. *Science*, 1960, 132, 805-806.
- WYRWICKA, WANDA, & DOTY, R. W. Electrical stimulation of "feeding centers." *Federat. Proc.*, 1962, 21, 353. (Abstract)
- WYRWICKA, WANDA, DOBRZECKA, C., & TARNECKI, R. On the instrumental conditioned reaction evoked by electrical stimulation of the hypothalamus. *Science*, 1959, 130, 336-337.

(Received June 29, 1962)

## EXPERIMENTAL INVESTIGATIONS OF THE GENESIS OF VISUAL SPACE PERCEPTION<sup>1</sup>

WILLIAM EPSTEIN

*University of Kansas*

The methodology and findings of 3 categories of experimental studies are reviewed and evaluated: (a) studies of organisms that have been deprived of visual stimulation until the time of testing, (b) studies of newborn organisms that are tested soon after birth, and (c) studies in which E controls the visual stimulation to which O is exposed from birth to the time of testing. The evidence provided by these experiments suggests that conclusions about the genetic basis of space perception will differ depending on the stimulus correlates under consideration.

This article is concerned with the genesis of space perception. It is necessary at the outset to distinguish the question under consideration from two other issues with which it is often confused. The first concerns the role of past experience in determining the current perceptual event. The second is related to the influence of specific practice on perceptual discrimination. Arguments intended to bear on the question of genesis are often derived from experiments dealing with the latter two issues. For example, it is asserted frequently that if perceptual judgments can be modified by certain types of previous experience, then experience must have been required for their original emergence. This argument has persisted since Helmholtz who wrote in 1866 that "whatever, therefore, can be overcome by factors of experience, we must consider as being itself the product of experience and training [Helmholtz, 1925, p. 13]." Despite the long history of this assumption it must be clear that evidence of the malleability of perception cannot entail any conclusion about the origin of the experience. Thus, one would hardly wish to conclude from the observation of the phenomenon of "memory color" (Dunker, 1939) that color perception is

learned. Pastore (1960) has called attention to a parallel example in the field of genetics where "evidence of environmental modification of a trait does not preclude its genetic determination [p. 94]." This argument is mentioned at this point only to make clear that the issue under consideration is *not* whether space perception can be modified by experience.

For this reason many investigations which might otherwise be included, for example, studies of adaptation to spatial displacement, were omitted (I. Kohler, 1951, 1962). This rule of omission is followed regarding the adaptation studies despite the fact that Kohler and others seem to feel that these experiments provide data relevant to a determination of the origins of space perception. Kohler has maintained that the conditions of displaced vision produced by the prisms reinstate in the adult an earlier state of perceptual development comparable to that of the infant. And that the processes that lead to adaptation to the experimentally induced displacement are reiterations of the processes of original perceptual development. A thorough review of the adaptation studies conducted by Stratton, Snyder and Pronko, Kohler, and others may be found in Smith and Smith (1962, Ch. 4-6).

The investigations reviewed in this paper were concerned mainly with the

<sup>1</sup> This article was written with the support of Research Grant MH 04153-C3 from the National Institute of Mental Health of the United States Public Health Service.

sufficient conditions for the *initial* occurrence of a response indicating that an organism is capable of spatial discrimination. Can these conditions be specified entirely in terms of the organism's native or congenital constitution or must a complete statement of these conditions include reference to processes of early learning?

### THREE METHODS OF INVESTIGATION

Until recently many American psychologists have been content to relegate the genetic question to the realm of philosophical inquiry. However, insofar as the question is concerned with the genesis of space perception and not with its epistemological status, it can only be decided empirically and is not properly a matter for philosophy. Fortunately, not all psychologists have abdicated their prerogatives of investigation and the last decade has witnessed a modest resurgence of interest in the problem. The experimental investigations can be classified usefully into three categories.

One category consists of studies of organisms that have been deprived of visual stimulation until the time of testing. On the first, or on a very early, occasion of stimulation their perception of space has been examined. When animals have been studied they have been reared in total darkness from birth to the time of testing. When human subjects have been studied, these have been (ideally) persons with congenital cataracts whose vision has been restored surgically.

The second category includes the investigations of naive, newborn organisms tested soon after birth. Animals are better suited to this approach than humans since human infants are capable of very few reliable overt behavioral responses soon after birth. Consequently, human infants have usually been studied at a later age (Fantz, 1961; Gibson & Walk, 1960). However, there

has been one study by Wertheimer (1961) of the auditory and visual space of a human neonate during the first 10 minutes of life.

The third category of experiments utilizes an entirely different approach. In these studies the experimenter controls the characteristics of the visual stimulation to which the animal is exposed from birth to the time of testing. The aim is to determine whether animals that have been subjected to different conditions of stimulation will respond differently when confronted with a relevant test discrimination. The rationale of this approach will be elaborated when Hess' (1950, 1956b) studies are considered.

### STUDIES OF VISUALLY DEPRIVED OBSERVERS

#### *Studies of Animals*

The earliest systematic study was performed by Lashley and Russell (1934). The object of the study was to determine whether rats reared in darkness from birth to 100 days of age would perceive variations of distance. The rats were induced to jump from one platform to a second platform whose distance was varied from 20 centimeters to 40 centimeters. An apparatus invented by Russell (1932) in order to study depth discrimination with normally reared rats was used. This apparatus allowed the experimenter to measure the force exerted by a rat in attempting to jump the given distance. Variations in force were interpreted as signifying corresponding variations in perceived distance. The experiment was conducted during a period of 3 days with the rats returned to the dark cage at the conclusion of each day's testing. On the first day the animal was removed from the dark cage and given five trials in *stepping* across a gap of 5 centimeters from the starting platform to the landing platform. Immediately following this

practice, the landing platform was moved to 20 centimeters for five trials. On the second day the animal began with five trials at 20 centimeters. This was followed immediately by a sequence of nine jumps in this order: three jumps at 40 centimeters, three jumps at 20 centimeters, three jumps at 40 centimeters. On the third day the rat's difference-threshold for distance was determined. For this purpose, each rat was given one trial at each of nine distances ranging from 24 to 40 centimeters in 2-centimeter intervals. The distances were presented in random order.

The results showed that upon first exposure the rats adjusted the force of their jumps to the variations of distance. Thus a rat that overestimated the 20-centimeter distance on the first test trial would consistently reduce the force of his jump during the next nine trials. However, the first time the 40-centimeter distance was presented (Trial 6 on the second day), the rat immediately reversed the trend and doubled the force. This occurred with every subsequent change of distance. On the tests for difference-thresholds the dark-reared rats were almost as accurate as light-reared rats that had been tested earlier by Russell (1932) in the same situation. Lashley and Russell (1934) concluded

that the visual perception of distance and gradation of force in jumping to compensate for distance are not acquired by learning, but are the product of some innately organized neural mechanism [p. 143].

Several doubts may be raised concerning the Lashley-Russell study. Greenhut and Young (1953) were unable to repeat Russell's (1932) original findings with 100-day-old and albino rats that had been reared normally. When they presented the initial test distances in random order they failed to find any correlation between the force of the jump and distance. However, it should be observed that Greenhut and Young used

electric shock in order to motivate the animals to jump. Despite some evidence to the contrary (Greenhut & Young, 1953, pp. 174-175), it is possible that this aversive stimulus disrupted the animal's spatiomotor coordination in many instances. In fact, Greenhut and Young (1953) reported that "the animals were emotionally disturbed throughout the experiment [p. 160]."

In addition to the results of Greenhut and Young, the Lashley-Russell experiment has been criticized on the grounds that opportunities for learning were not completely eliminated. In particular it has been noted that on every second day the dark box was opened in dim light while food was supplied. Although each interval of light stimulation did not exceed 10 seconds, it is possible that the cumulative exposure was sufficient to produce substantial learning. The five practice trials which preceded the first test jump may have provided an opportunity for rapid learning. This argument gains plausibility from the general observation that organisms which have reached an advanced maturational level require less practice to achieve a specified performance level than organisms that begin their practice at an earlier level of maturation (Munn, 1955).

Lashley and Russell (1934) recognized the problems discussed above and presented counterarguments that are persuasive. However, the only effective way to deal with these objections is to perform new experiments which are more successful in eliminating opportunities for learning. The main flaw in the Lashley-Russell study stemmed from the nature of the response that they studied. The jumping response is a complex motor skill which requires practice. Rats reared in the dark are deficient in this skill. Therefore, it was necessary to introduce the practice trials, thus providing an opportunity for both visual and

motor learning. A response indicator which is part of the organism's unlearned response repertoire would eliminate this necessity. Since the performance of the response *per se* would require no practice, the subject could be tested on the very first occasion of his exposure to the stimulus without prior practice of the response.

Kurke (1955), using a technique similar to that employed in the earlier studies by Spalding (1875) and Thorndike (1899), studied the depth perception of chicks reared under different conditions of visual and motor experience. Kurke's procedure was based on the observation that lone chicks will run to join their fellows. The chick was placed on a platform whose height could be adjusted. From the platform the subject could see a group of six chicks. The assumption was made that if the chicks perceived the height of the platform, then they would be reluctant to leave the platform before it was lowered to a safe level. The platform was set initially at a height of 21 inches, and if the chick did not jump within 30 seconds, the platform was lowered 3 inches. This was repeated until all subjects had jumped or the platform was on ground level. Dark-reared chicks were tested at 1, 2, or 3 days of age. Their performance was compared with a group of 3-day-old control chicks raised in normal lighting. The very young chicks (1 and 2 days) refused to leave the platform regardless of its height. The 3-day-old dark-reared chicks did not differ significantly from the light-reared control chicks. Nor did they differ significantly from 10-day-old light-reared chicks whose opportunities for jumping experience was limited. However, 10-day-old light-reared chicks which were subjected to enforced jumping activity differed from the other groups in that they jumped from the platform at a greater height. Kurke (1955) concluded:

that the lack of visual experience does not appreciably affect the chick's perception of depth at an early age. However, the further development of depth discrimination apparently depends in some manner on the integration of experientially determined kinesthetic cues [p. 195].

Kurke's experiment is not very convincing. As was the case with the Lashley-Russell study, the difficulty is due mainly to the response measure. The response of jumping is determined by other variables besides perceived height. Among these is the level of muscular coordination and jumping skill which has been achieved. This leads to ambiguities in interpreting the results. Differences in performance may reflect differences in depth discrimination, but it could also be the case that subjects who are equally able to discriminate depth will vary in performance because of different levels of motor achievement.

The resolution of this methodological problem has been accomplished in several studies of color vision (Hess, 1956a), form discrimination (Fantz, 1957), and recently in a study of depth discrimination by dark-hatched chicks (Fantz, 1958a). The chick's discrimination of depth or solidity was examined by observing the distribution of unreinforced pecks in response to the simultaneous presentation of a spherical and plane circular surface. The rationale underlying the test is that if the chick shows a significant preference for one of the stimuli, that is, pecks more frequently at one, then the chick must be making perceptual discriminations between them. If the stimuli differ only on the relevant dimension, for example, solidity, then the subject must be able to discriminate along this dimension. It should be noted that, in addition to providing a resolution of the methodological difficulties described above, the stimulus-preference method also reduces the likelihood of confusing two different achievements: (a) learning a discrimina-

tion between stimuli to some arbitrarily assigned experimental criterion, and (b) discriminating between stimuli. While the former achievement presupposes the latter, it cannot be utilized as an index of the latter achievement except with certain qualifications. Suppose the experimenter confronts the subject with the task of learning a discriminatory positive response to a convexity as opposed to a concavity. As a measure of learning the experimenter uses the criterial level of nine correct responses in a block of 10 successive trials. This accomplishment presupposes that the subject can perceive the depth difference between the two stimuli, although he must learn that this difference is the criterial attribute. Therefore, it would be erroneous to conclude that a subject that fails to achieve the criterial level, for example, correctly selects the convexity in only 8 of 10 trials, cannot discriminate depth. It is equally plausible to infer that depth is discriminated, but has not been selected unequivocally by the subject as the basis for responding. This argument has been made earlier by Zuckerman and Rock (1957) in their analysis of the genesis of form perception.

Fantz (1958a) hatched 94 chicks in complete darkness and tested them at the moment of their initial visual stimulation. The chicks had no experience with food or water prior to or during the experiment. In the first experiment chicks were tested singly, and in the second experiment they were tested in pairs. The test consisted of a 5-minute exposure to a pair of stimuli. The stimuli were two hemispheres, one presented so that the curved surface faced the chicks and the other presented so that only a flat circle was visible. The stimuli were presented either under direct lighting which produced strong shading or under diffuse lighting which reduced the brightness gradients greatly.

Under both conditions the curved surfaces received significantly more pecks than the plane surfaces. Apparently these chicks, with no visual experience, discriminated between the solid and plane stimuli. Comparable findings were reported by Fantz (1957, Experiment 5) in his studies of form preference in newly hatched chicks. While these newborn chicks seemed to make no use of light and shade distribution in discriminating depth, Fantz reported that slightly older chicks that had visual experience showed an enhanced preference for spherical surfaces when shading was present. This observation is consistent with Hess' (1950) findings which will be described later.

Walk and Gibson (1961) realized the same methodological objectives by using the visual cliff apparatus. This is a simulated cliff consisting of a board laid across a sheet of heavy glass which is supported at a desired height above the floor. On one side of the board a sheet of patterned material, for example, checkerboard pattern, is placed flush against the undersurface of the glass. On the other side the same patterned material is laid on the floor below the glass. This is the cliff side. If an organism which discriminates depth is placed on the center board it will avoid the cliff side and step on to the shallow side. If the organism cannot distinguish between the shallow side and the cliff side, then no preference should be observed. In these experiments, as was the case in Fantz' study, the discriminatory response required no pretraining in light, and the discrimination did not entail cognitive, problem solving elements. Rats were dark-reared and tested at the age of 90 days. These rats showed the same preference for the shallow side as a group of 90-day-old light-reared rats.

Walk and Gibson's results for the dark-reared rats were confirmed in a modified experiment by Nealey and Ed-

wards (1960). Nealey and Edwards noted that the dark-reared rats in the Walk-Gibson experiment were subjected to a light-adaptation period of 20 minutes prior to testing. This adaptation period may have provided the opportunity for learning to occur. For this reason Nealey and Edwards repeated the Walk-Gibson study with the following modification. During the 20-minute period subjects were kept in a detention box which was so contrived as to permit light adaptation without pattern vision. These rats showed the same preference for the shallow side as rats that received the Walk and Gibson treatment. Nealey and Edwards also tested 24 enucleated rats in order to determine whether a preference could be obtained on other than visual bases. These rats did not show a reliable preference.

In further experiments Walk and Gibson (1961) tested dark-reared rats when the available stimuli for depth discrimination were reduced. When motion parallax was eliminated, leaving only the difference in density of optical texture as a stimulus for depth discrimination, no preference for the shallow side was exhibited. However, when parallax was retained, in the absence of the textural difference, the rats showed a strong avoidance of the cliff side. This is interpreted by Walk and Gibson to indicate that only motion parallax is an innate cue for depth discrimination on the visual cliff. Since visually experienced animals show a preference for the shallow side when cued only by textural density while dark-reared rats show no preference, the cue of textural density must be learned. This conclusion is reasonable, but it is not the only interpretation which can be given to these findings. An alternative stems from the observation that the procedures did *not* in fact isolate motion parallax from textural density. Instead, the two gradients were opposed to each other. One gradi-

ent was fixed to produce equidistance while the other was fixed to produce differential distance. In this conflict situation motion parallax was decisive. However, it cannot be concluded that the effectiveness of textural density requires learning. This conclusion can be derived only from the observation of performance in response to textural density in isolation.

Regardless of the above qualification, the general trend of the evidence obtained in the visual cliff experiments is clear. The evidence led Gibson and Walk (1960) in an earlier article "to venture the rather broad conclusion that a seeing animal will be able to discriminate depth when its locomotion is adequate, even when its locomotion begins at birth [p. 71]."

### *Studies of Humans*

The case of the congenitally blind person whose vision is restored surgically would appear to provide an opportunity for gaining decisive evidence concerning the origin of visual perception. Since Molyneux's famous inquiry addressed to Locke, philosophers and psychologists have looked to these cases for evidence. Several reviews have appeared which evaluate the results of the investigations of restored vision (Dennis, 1934; Senden, 1932; Wertheimer, 1951). Prior to evaluating the evidence it will be useful to consider several relevant methodological and definitional questions.

1. The perception of space includes several related, but distinguishable and perhaps functionally independent, experiences. There is the fundamental and relatively undifferentiated experience of the extensity of the visual scene. The scene appears to extend in the third dimension and not merely to be a plane surface like a painting in the frontal-parallel plane. Within the scene, objects have a specific location relative to us (absolute distance) and also a location relative to other objects in the scene

(relative distance). In addition, each object will appear to be some linear size, either absolutely or relative to other objects. Finally, some objects will appear to be solid and others two-dimensional. Every student of perception knows that these aspects of the visual world are closely related. Nevertheless, it does not follow that an observer whose vision is tested immediately after surgical restoration and who is found deficient in one of these experiences is totally lacking in space perception. Therefore, broad conclusions concerning space perception based on performances related to only one aspect of space perception should be considered cautiously.

2. These remarks lead directly to a consideration of the method of testing. It is easier to specify prohibitions than to advance positive proposals. Tasks which demand prior visual learning for successful execution are unacceptable. One illustration of this is the designation of an object by the appropriate verbal label. Thus the subject might be confronted with a number of objects and the question "Which of the objects is a cube?" This question cannot be answered without a prior association of the auditory stimulus, "cube," with the *visual* stimulus produced by a cube. Obviously an observer with no prior visual experience will lack this association. A better procedure would be to confront the subject with equivalent plane and solid objects and then simply ask the subject to report whether he perceives several different objects or several instances of the same object.

Similar reservations exist regarding another testing method which has often been used. Investigators have sought to determine whether apparent size is invariant with changes in distance for the subject whose sight is newly restored. The presence or absence of constancy is used to derive conclusions concerning distance perception. These conclusions

are open to question. Size judgment is a complex performance which is regulated by many variables in addition to distance perception. One powerful determinant is the conceptualization of the relationship between size and distance which the subject brings to bear on the experimental task (cf. Carlson, 1960, 1962). Subjects with normal vision will produce results ranging from marked underestimation to marked overestimation of size depending on their attitude of observation. In the absence of knowledge about the subject's attitude of judgment, it is impossible to interpret the results of his performance. Therefore, prior information concerning the preoperative conceptualization of the visual world is a prerequisite for interpreting the performance of the newly-seeing subject.

The question with which this discussion was started remains. How should the subjects be tested? No ordered schedule of tests will be introduced here. However, it seems best that an open-ended question should initiate the inquiry. It should not direct the subject to any specific aspect of the visual world. A simple question such as "What do you see?" would meet this requirement. The response to this question might introduce the experimenter to the subject's preferred language of perceptual experience. Later questions or instructions could then be phrased in the subject's language so that there would be assurance of their meaningfulness to him.

3. Senden (1932), Dennis (1934), and Wertheimer (1951) have described the postoperative visual disabilities which interfere with good vision. For instance many patients suffer from exaggerated and involuntary nystagmus which may continue for several weeks. This interferes with the normal functioning of accommodation and convergence. Two additional postoperative effects which occur frequently are acute dazzle and a narrowing of the visual

field. Obviously, tests which produce negative results are equivocal if the results are obtained prior to the adjustment of the visual mechanisms. On the other hand, if the experimenter elects to postpone the test for several weeks to allow the inhibitory effects to diminish, then there is a possibility that some learning occurred during the delay.

A review of the literature is disappointing. None of the subjects have been tested adequately (for present purposes), and most of the cases have been described only cursorily. The chief source of information is Part III of Senden's (1932) book. Senden reviews the evidence regarding the perception of spatial extension per se, relative distance, and solidity.

The reports concerning spatial extensity are unanimous, leading Senden to the conclusion that the newly-seeing have an immediate impression of depth. The objects in the field appear to be spatially separated from the subject. The patient localizes objects at certain indefinite distances although he cannot estimate their absolute distance. Thus the patient will reach out to grasp an object which is several yards away, or he will overreach an object which is only 1 foot from him. However, he *does* reach out into space and does *not* localize the object at the plane of the eye. This reaching response is qualitatively different from the normal groping that the blind usually execute in order to identify objects tactually.

The evidence concerning relative distance is ambiguous. The most frequently used procedure entailed a crude test of size constancy. Most examiners report that constancy is absent. This observation suggests that differences in distance were not perceived. However, as was noted earlier, this conclusion can be challenged.

Most of the cases reviewed by Senden (1932, pp. 264-271) indicate that the

discrimination of solidity is absent. For example one subject was unable to distinguish a ping-pong ball from a comparable plane figure, for example, a white disc of equal diameter. For another subject a table in the center of the room appeared flat against the farther wall.

Recently London (1960) provided a summary of a Russian report on the postoperative newly-seeing which was published originally in 1953. This report is of special interest in that, unlike the Senden report, the writer, Pokrovskii, performed the surgery and the postoperative visual tests. Of the six cases described only two are directly pertinent. These are cases of two children, born with mild cataracts which allowed the patients the experience of light although pattern vision was not possible. London presents only the author's statement of results omitting details of the testing procedure. In the only statement dealing with space perception Pokrovskii reported that

the children were unable by vision alone to determine distance or, more exactly, the distance of the nearest objects. When walking, they collided with these objects [London, 1960, p. 479].

This review of the investigation of the newly-seeing has revealed the unsatisfactory status of this evidence. The techniques of investigation are crude and the data are incomplete. In fact, there is not a single systematic parametric study of the visual perception of the newly-seeing. In the absence of more acceptable studies, the most prudent course is to refrain from introducing the findings described above as evidence. It is only fair to add that this judgment is not shared by all who have considered these investigations. In his early theoretical statements Hebb (1949) assigned great significance to Senden's findings and he has more recently reaffirmed his evaluation (Hebb, 1963).

## STUDIES OF VISUALLY NAIVE OBSERVERS

The study of the newborn organism requires that a reliable unlearned indicator response be available immediately after birth. This requirement can be satisfied for many organisms although not in all testing situations. For example, the human neonate cannot be tested with the visual cliff apparatus. In order to employ the visual cliff, the experimenter must await the development of locomotion at which time the organism is no longer perceptually naive. Nevertheless, the human infant is not entirely excluded from investigation. Fantz (1961) and Wertheimer (1961) have studied the human neonate by observing the systematic changes in orientation of the eyes in response to various stimuli.

*Studies of Animals*

Walk and Gibson's (1961) work with the visual cliff apparatus did not include testing newborn observers. However, several species (e.g., the chick and the goat) are able to locomote during the first day of life. Gibson and Walk reported that none of their 1-day-old chicks, goats, or lambs ever stepped onto the glass on the cliff side. Gibson and Walk consider it very likely that similar results would be obtained with newborn animals as well. This expectation received support from Tallarico's (1961) study of the choice behavior of 3-hour-old chicks on the visual cliff. Of the 320 chicks observed, 90.6% chose the shallow side. Those chicks that stepped onto the glass on the cliff side seemed to do so inadvertently.

In another study Fishman and Tallarico (1961) also studied visual depth perception in 3-hour-old chicks. As an indicator response Fishman and Tallarico selected the avoidance response made on the first occasion of an approaching object. Six groups of chicks were tested. Three groups consisted of chicks reared in the light. The re-

maining three groups consisted of chicks reared in the dark. All chicks were tested individually at the age of 3 hours. There were three conditions of testing. In Condition FP (feigned poke) one experimenter held the chick in his fist while the second experimenter feigned a poke at the left eye of the chick. This consisted of moving a black pocketcomb from a distance of 1 foot to within  $\frac{1}{2}$  inch from the subject's eye at the rate of 1 foot per second. The experimenter recorded the occurrence of a spontaneous head movement away from the direction of the poke. In Condition EF (extended fan) no poke was feigned. Instead a black fan held about 1 foot from the subject was opened in about a second at a right angle to the subject's line of sight. In Condition C (control) the experimenter held the chick as before and for the same length of time but without introducing any stimulus.

The dark-reared chicks did not differ from the light-reared chicks under any of the three conditions. The control chicks made very few "avoidance" head movements. Of the 73 control chicks only five spontaneous head movements were observed which were similar to the aversive movements of the experimental subjects. For both light-reared and dark-reared chicks the number of avoidance responses under Condition FP were significantly greater than under Condition EF. In fact only 1 of 18 chicks gave an avoidance response to EF, while 44 of 96 chicks responded to FP. These results are evidence that visually naive chicks can discriminate depth.<sup>2</sup> The

<sup>2</sup> However, Riesen (1947, 1950) reported that dark-reared chimpanzees "did not blink at a threatening motion toward the face [and] when an object was advanced slowly toward the face there was no reaction [1950, p. 17]." Riesen's work has not been described in detail because it did not include any systematic tests of the subject's depth discrimination. In addition his chimps suffered from considerable oculomotor disturbance at the time of testing.

avoidance reaction seems to be a response to the unique stimuli provided by an approaching object. It is not a startle response made to a novel stimulus since it did not occur under Condition EF. Nor does it occur with great frequency in the absence of stimulation (Condition C).

It would be interesting to determine whether the newborn chick can indeed distinguish between the optical stimuli produced by a change in position and the comparable stimuli produced by an equivalent change in the size of a stationary object. Apparently infant rhesus monkeys respond to the isolated presentation of a closed contour undergoing symmetrical expansion as if it were an approaching object (Schiff, Caviness, & Gibson, 1962). On the other hand the chicks in Fishman and Tallarico's study did distinguish between the optical stimuli produced by FP and EF. However, since these two conditions were not exactly comparable no firm conclusion can be drawn concerning our question. The apparatus and stimulus events employed by Schiff et al. would be very appropriate for this purpose.

#### *Studies of Humans*

There have been even fewer studies of naive humans. Although the infants studied by Fantz (1961) had a median age of 15 weeks this investigation will be included. The only experiment which investigated newborn humans was reported by Wertheimer (1961).

Fantz (1961) employed the stimulus-preference method using the direction of orientation of the eyes as an indicator of preference. A solid object and a comparable flat object were exhibited for a series of 20-second trials and the subject's direction of gaze was recorded. The criterion of fixation was the superposition of the reflection of the object over the pupil. The subjects were 52 infants with a median age of 15 weeks.

The test objects were a sphere and a disc of equal diameter. Each infant was tested under eight conditions: binocular and monocular vision, presence and absence of surface texture, direct and diffuse lighting. A differential response was obtained only in the tests using textured objects under direct lighting. The other six conditions did not yield a differential response. A curious finding emerged from these results. Infants under 3 months did not give a differential response in the binocular test but the same subjects showed a clear preference in the monocular test. This result, together with others, suggested to Fantz (1961)

that the use of both eyes interferes with vision in the early months before the development of good binocular coordination, while binocular vision improves visual performance later on [pp. 29-30].

The results led Fantz to conclude that young infants discriminate solidity when gradients of texture and brightness are available. Binocular cues did not appear to be essential in the presence of these gradients nor were binocular cues sufficient in the absence of differential gradients. It is interesting to note that whenever a differential response occurred it was a preference for the solid object. It may be that the basis for this preference is the more general visual interest which young infants display in response to complex visual patterns (Berlyne, 1958; Fantz, 1958b). It is immediately evident that the presence of a general tendency of this sort introduces difficulties in interpreting Fantz' (1961) results. It is not possible to know whether the differential responsiveness to the sphere reflects a depth discrimination or a pattern preference.

Wertheimer (1961) studied the psychomotor coordination of auditory and visual space of one human infant before she was 10 minutes old. The subject was born without anesthesia, by natural

childbirth. Testing commenced 3 minutes after birth. On each trial a click was sounded next to the right or left ear of the subject who was lying on her back. Two experimenters recorded the infant's eye movements in response to the click. Fifty-two successive trials were undertaken. On 45 of the 52 trials both experimenters were in agreement concerning the eye movement. In 23 of these cases there was no movement. On the remaining 22 trials 18 of the eye-movement responses were in the direction of the click and four in the opposite direction. Thus within 10 minutes after birth auditory localization and a coordinate primitive spatial localization is present in the human infant.

#### EXPERIMENTAL VARIATION OF THE RELEVANT STIMULATION

The only investigations in this category are those performed by Hess (1950, 1956b, 1961). The earlier study was concerned with the development of depth responses to shading cues. Hess (1950, 1961) reared two groups of 50 chicks under different conditions of illumination. The control chicks were reared in cages illuminated from above. The experimental chicks were reared in cages which were illuminated from below. Thus the two groups experienced different distributions of light and shade. They remained in these cages for a period of 7 weeks after hatching.

The test enclosure was a circular cage illuminated by lights at opposite sides of the enclosure. The test objects were two photographs placed side by side. The photographs showed grains of wheat in their natural coloring but about twice the normal size. One picture was right-side up and the other upside down. In the first experiment chicks were tested at the age of 7 weeks. They were put into the test cage in groups of two to four chicks. First they were permitted to

eat a few grains which were fastened vertically on a cardboard. Then the photographic plate was substituted and the chicks' responses to the photographs were recorded for a 10-minute interval. The chicks were then returned to their rearing environments and retested a week later. At that time the illumination in the test cage originated from the direction opposite to that of the light in the rearing cages. Because of the wide range in the number of responses made by individual chicks Hess analyzed the first pecks only. The results were the same for the test and retest. The chicks showed a statistically significant tendency to peck at the photograph representing the distribution of shading consistent with their rearing. For instance, on the retest, 93% of the first pecks made by the control group were made on the photograph showing grain illuminated from above, while only 6.1% of the first pecks of the experimental groups were on that photograph.

The second experiment studied the development of depth responses in the experimental animals. At 1-week intervals beginning at 1 week and continuing to 7 weeks of age, groups of chicks were taken from the rearing cage and tested. The main finding of the experiment was that the establishment of a significant preference for the photograph representing overhead illumination occurred earlier in the control than the establishment of a significant preference for the opposite photograph in the experimental group.

The results of the first experiment show that the chick's responses to light and shade cues can be governed by the direction of lighting in its environment. This does not require the conclusion that no preference would be exhibited in the absence of experience. It may be that an unlearned response to illumination from above exists, but that this innate

preference can be extinguished if the environment does not support it. In fact, such may be the implication of the results of the second experiment. If no initial unlearned preference existed then the control and experimental groups should develop a preference with equal rapidity. On the other hand, if the experimental group has an initial preference for illumination from above, then they must unlearn this preference prior to acquiring the new response. This could explain the finding that the experimental group established a significant preference later than the control group.

One incidental finding of this study which warrants notice is that the photographs of grain were responded to as if they were real grains. Apparently, for chicks, pictorial representations yield depth responses in the absence of a history of experience with representations. Although it has been amply demonstrated that human observers may have compelling experiences of depth when presented with photographs (e.g., Smith & Smith, 1961), it is not known whether experience with photographic representations is a necessary condition. A recent study by Hochberg and Brooks (1962) found that a child "is capable of recognizing pictorial representations of solid objects without specific training or instruction [p. 628]." They infer, therefore, "that there must be an unlearned propensity to respond to certain formal features of lines-on-paper [p. 628]." This does not answer our question; however, it does lead to the expectation that similar results will be obtained for depth perception.

In a later study Hess (1956b) examined a more general question: "When an organism first perceives the environment can it accurately see where things are [p. 74]?" Specifically, the experiment attempted to determine whether

the chick's ability to localize and peck accurately at objects is innate or learned. Chicks were hatched in darkness and divided into two experimental groups and one control group. The experimental chicks were fitted with hoods containing prisms which displaced the visual field either 7 degrees to the right or 7 degrees to the left. The control chicks were fitted with hoods which did not contain image-displacing lenses. When they were 1 day old the chicks were tested for pecking accuracy. They were allowed to peck at a small brass nail embedded in modeling clay. The clay provided a visual record of the location of the chick's pecks.

The following is the rationale of this procedure. If spatial localization is learned then the early responses of the chick should be randomly distributed in the visual field. Only through practice should the pecks converge upon and finally strike the target. If this is a valid analysis of the development of localization, then the displacing prisms should not affect the development of localization. In terms of this analysis the experimental and control groups both must learn to localize objects through a gradual process of trial and error pecking. On the other hand, if spatial localization is innate then the first pecks of the experimental chicks should cluster about a point 7 degrees to the right or left of the target while the pecks of the control chicks should center about the target. The results agreed with the latter expectations.

Following the test on the first day Hess divided each of the three groups in half. Half of each group was placed in a living enclosure in which grain was scattered loosely on the floor. The other half of each group was placed in a box which contained bowls of mash. When the chicks were 3-4 days old, they were tested again. There was no difference

between the two halves of the control group. Both halves showed an increase in accuracy and both were in good physical condition. The two halves of the experimental group also improved their accuracy of pecking. However, their improvement was represented by a tighter cluster 7 degrees to the right or left of the target. The two halves of the experimental groups differed markedly in physical health. The group whose living cage contained bowls of mash was as healthy as the control group. The half whose living cage contained scattered grain only were in poor physical condition. Two of these chicks died on the next day. Hess concluded that spatial localization is innate for the chick and that the chick is unable to unlearn the innate localization of objects even when survival is at stake. A similar absence of adaptation is observed in the persistent spatial disorientation which results from the more dramatic modifications of sensory input produced by surgical procedures (Sperry, 1951, 1958). For example, surgical rotation of the eyes, producing an inverted visual field, leads to mislocalization without any subsequent adaptation.

### CONCLUSION

No simple decision about the genesis of space perception is possible. Nor does it seem likely that a conclusion of general applicability is forthcoming. Instead the trend of the evidence strongly suggests that separate statements will have to be made about each of the various stimulus correlates of depth perception. For example, there is evidence that motion parallax is an unlearned correlate of relative distance while differences in textural density are learned correlates (Walk & Gibson, 1961). This conclusion was clearly anticipated by Carr (1935) in his treatise on space perception:

There is no single phenomenon or problem of space. It is impossible to make any statement as to the native or empirical nature of space in general [p. 5; italics added].

This conclusion poses two questions whose resolution would provide further clarification of the issue. The first concerns the modifiability of innately determined perceptual preferences by later experience. It seems likely that the various stimulus correlates of depth perception will vary in their susceptibility to the modifying influence of special experience. The second question concerns the nature of the learning processes which underlie the formation of associations between specific cues and the concomitant visual experience of depth. This latter question is an extremely difficult one as Pastore (1956) has pointed out in his critique of the empiristic viewpoint. Nevertheless, as evidence is obtained (e.g., Hess, 1950) that certain cues are learned, the need for an analysis of this question becomes compelling.

### REFERENCES

- BERLYNE, D. E. The influence of albedo and complexity of stimuli on visual fixation in the human infant. *Brit. J. Psychol.*, 1958, 49, 315-318.
- CARLSON, V. R. Overestimation in size-constancy judgments. *Amer. J. Psychol.*, 1960, 73, 199-213.
- CARLSON, V. R. Size-constancy judgments and perceptual compromise. *J. exp. Psychol.*, 1962, 63, 68-73.
- CARR, H. *An introduction to space perception*. New York: Longmans, Green, 1935.
- DENNIS, W. Congenital cataract. *J. genet. Psychol.*, 1934, 44, 340-350.
- DUNCKER, K. The influence of past experience upon perceptual properties. *Amer. J. Psychol.*, 1939, 52, 255-265.
- FANTZ, R. L. Form preferences in the newly-hatched chick. *J. comp. physiol. Psychol.*, 1957, 50, 422-430.
- FANTZ, R. L. Depth discrimination in dark-hatched chicks. *Percept. mot. Skills*, 1958, 8, 47-50. (a)
- FANTZ, R. L. Pattern vision in young infants. *Psychol. Rec.*, 1958, 8, 43-47. (b)
- FANTZ, R. L. A method for studying depth

- perception in infants under six months of age. *Psychol. Rev.*, 1961, 11, 27-32.
- FISHMAN, R., & TALLARICO, R. B. Studies of visual depth perception: II. Avoidance reaction as an indicator response in chicks. *Percept. mot. Skills*, 1961, 12, 251-257.
- GIBSON, ELEANOR J., & WALK, R. D. The "visual cliff." *Scient. American*, 1960, 202, 64-71.
- GREENHUT, ANN M., & YOUNG, F. A. Visual depth perception in the rat. *J. genet. Psychol.*, 1953, 82, 155-182.
- HEBB, D. O. *Organization of behavior: A neurological theory*. New York: Wiley, 1949.
- HEBB, D. O. The semiautonomous process: Its nature and nurture. *Amer. Psychologist*, 1963, 18, 16-27.
- HELMHOLTZ, H. v. *Physiological optics*. (Ed. by J. P. S. Southall) Vol. 3. New York: Optical Society of America, 1925.
- HESS, E. H. Development of the chick's responses to light and shade cues of depth. *J. comp. physiol. Psychol.*, 1950, 43, 112-122.
- HESS, E. H. Natural preferences of chicks and ducklings for objects of different colors. *Psychol. Rep.*, 1956, 2, 477-483. (a)
- HESS, E. H. Space perception in the chick. *Scient. American*, 1956, 195, 71-80. (b)
- HESS, E. H. Shadows and depth perception. *Scient. American*, 1961, 204, 138-148.
- HOCHBERG, J., & BROOKS, VIRGINIA. Pictorial recognition as an unlearned ability: A study of one child's performance. *Amer. J. Psychol.*, 1962, 75, 624-628.
- KOHLER, I. Über aufbau und Wandlungen der Wahrnehmungswelt. *Oesterr. Akad. Wiss Philos. Histor. Kl. Switz.-Ber.*, 1951, 227, 1-118.
- KOHLER, I. Experiments with goggles. *Scient. American*, 1962, 206, 62-84.
- KURKE, M. I. The role of motor experience in the visual discrimination of depth in the chick. *J. genet. Psychol.*, 1955, 86, 191-196.
- LASHLEY, K. S., & RUSSELL, J. T. The mechanism of vision: A preliminary test of innate organization. *J. genet. Psychol.*, 1934, 45, 136-144.
- LONDON, I. A Russian report on the post-operative newly seeing. *Amer. J. Psychol.*, 1960, 73, 478-482.
- MUNN, N. L. *The evolution and growth of human behavior*. New York: Houghton Mifflin, 1955.
- NEALEY, S. M., & EDWARDS, BARBARA J. "Depth perception" in rats without pattern vision experience. *J. comp. physiol. Psychol.*, 1960, 53, 468-469.
- PASTORE, N. An examination of one aspect of the thesis that perception is learned. *Psychol. Rev.*, 1956, 63, 309-316.
- PASTORE, N. Perceiving as innately determined. *J. genet. Psychol.*, 1960, 96, 93-99.
- RIESEN, A. The development of visual perception in man and chimpanzee. *Science*, 1947, 106, 107-108.
- RIESEN, A. Arrested vision. *Scient. American*, 1950, 183, 16-19.
- RUSSELL, J. T. Depth discrimination in the rat. *J. genet. Psychol.*, 1932, 4, 136-159.
- SCHIFF, W., CAVINESS, J. A., & GIBSON, J. J. Persistent fear responses in rhesus monkeys to the optical stimulus of looming. *Science*, 1962, 136, 982-983.
- SENDEN, M. v. *Raum und Gestaltauffassung bei operierten Blindgeborenen vor und nach der Operation*. Leipzig, Germany: Barth, 1932.
- SMITH, K. V., & SMITH, W. M. *Perception and motion*. Philadelphia, Pa.: Saunders, 1962.
- SMITH, PATRICIA C., & SMITH, O. W. Ball throwing responses to photographically portrayed targets. *J. exp. Psychol.*, 1961, 62, 223-233.
- SPALDING, D. A. Instinct and acquisition. *Nature, London*, 1875, 12, 507-508.
- SPERRY, R. W. Mechanisms of neural maturation. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
- SPERRY, R. W. Physiological plasticity and the brain circuit theory. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958.
- TALLARICO, R. B. Studies of visual depth perception: III. Choice behavior of newly hatched chicks on a visual cliff. *Percept. mot. Skills*, 1961, 12, 259-262.
- THORNDIKE, E. L. The instinctive reactions of young chicks. *Psychol. Rev.*, 1899, 6, 282-291.
- WALK, R. D., & GIBSON, ELEANOR, J. A comparative and analytical study of visual depth perception. *Psychol. Monogr.*, 1961, 75(15, Whole No. 519).
- WERTHEIMER, M. Hebb and Senden on the role of learning in perception. *Amer. J. Psychol.*, 1951, 64, 133-137.
- WERTHEIMER, M. Psychomotor coordination of auditory and visual space at birth. *Science*, 1961, 134, 1692.
- ZUCKERMAN, C. B., & ROCK, I. A reappraisal of the roles of past experience and innate organizing processes in visual perception. *Psychol. Bull.*, 1957, 54, 269-296.

(Received August 14, 1962)

# ANXIETY (DRIVE) LEVEL AND PERFORMANCE IN EYELID CONDITIONING<sup>1</sup>

KENNETH W. SPENCE

University of Iowa

Studies from the Iowa laboratory and elsewhere that have involved a comparison of the eyelid conditioning performance of Ss scoring at the extremes of the Taylor Manifest Anxiety (MA) scale are reviewed. In 21 of 25 independent comparisons, differences between groups were in favor of the high anxiety (HA) Ss, with the majority being statistically significant. Although these data provide substantial confirmation of the implication of the drive interpretation of MA scale that HA Ss should exhibit a higher level of performance than LA Ss, an attempt was made to ascertain what factors might be responsible for failure of the difference to occur in some studies. The major factors appeared to be small numbers of Ss and the presence of "voluntary form" responders in the samples. Significant differences appear to be related to the degree of experimental naiveté of the Ss and the extent to which the experimental situation is designed to arouse some degree of apprehensiveness.

Recently King, Kimble, Gorman, and King (1961) reported failure to find a significant relation between level of emotionality as measured by the Taylor Manifest Anxiety (MA) scale and performance in aversive (eyelid) conditioning. In the light of the previous evidence, this finding was rather surprising and the authors were able to conclude only that emotionality or anxiety must be an interacting variable that is related to conditioning under certain as yet unknown conditions and not under others. The purpose of the present article is (a) to review the data bearing on the relation between the MA scale and aversive conditioning, only a small fraction of which was cited by King et al., and (b) to examine this evidence in detail in an attempt to suggest what the factors are that might be responsible for the failure of the relationship to appear in some studies.

In presenting these data, the results

of studies comparing performance in simple conditioning of Ss preselected on the basis of extreme MA scale scores will first be summarized. The second group of studies to be reviewed also involves comparison of the performance of preselected Ss but in differential conditioning. Thirdly, the results of a number of previously unpublished comparisons based on simple conditioning data from studies conducted in the Iowa laboratory for other purposes will be presented. The Ss serving in these latter studies were unselected with respect to the MA scale. Since their scores on the test were available, it was possible to determine the differences in the conditioning performance between Ss who scored at the extremes of the scale.

## STUDIES INVOLVING PRESELECTED EXTREME GROUPS

*Simple Conditioning.* The findings of the previously published Iowa studies and one unpublished study directly concerned with comparing the conditioning performance of preselected high and low anxiety (HA and LA) Ss are summarized in Table 1. The criteria of

<sup>1</sup> This study was carried out as part of a project concerned with the influence of motivation on performance in learning under Contract Nonr-1509(04), Project NR 154-107 between the University of Iowa and the Office of Naval Research.

TABLE 1

IOWA STUDIES OF CONDITIONING PERFORMANCE OF PRESELECTED HA AND LA SCALE Ss

Iowa experiments	No. trials	Ready signal	UCS (psi)	No. Ss	Percent CR		Difference (H-L)	<i>p</i>
					HA	LA		
1. Taylor (1951)	80	Yes	1.6	60	59.6	27.9	21.7	.001
2. Spence-Taylor (1951)	100	Yes	.6	50	48.2	33.8	14.4	.05 <sup>a</sup>
	100	Yes	2.0	50	55.0	41.7	13.3	.05
3. Spence-Farber (1953)	60	Yes	1.0	64	48.8	34.1	14.7	.05
4. Spence-Beecroft (1954)	50	Yes	1.0	45	56.5	36.3	20.2	.02
5. Spence-Weyant (1960)	100	No	.25	36	41.8	28.6	13.2	.10 <sup>b</sup>
	100	No	2.0	36	65.4	53.2	12.2	.10
6. Spence (unpublished)	80	Yes	.25	60	36.5	21.6	14.9	.02 <sup>c</sup>
	80	Yes	1.50	60	48.0	38.8	9.2	.02

<sup>a</sup> The *F* value based on all groups provided a *p* value < .01.<sup>b</sup> The *F* value based on all groups provided a *p* value < .05.<sup>c</sup> The *F* value based on all groups provided a *p* value < .01.

selection of Ss in these studies varied slightly, ranging from raw scores of 20 to 24 as the lower limit of the HA Ss and from 7 to 9 as the upper limit of the LA Ss. These scores roughly mark the upper and lower twentieth to twenty-fifth percentiles of the distribution of *MA* scale scores made by students in the introductory course in psychology at Iowa. Shown in the successive columns of the table are: (a) the reference studies, (b) the number of trials over which the conditioning measures were obtained, (c) the use or nonuse of a ready signal, (d) the strength of the UCS (air puff) employed, (e) the total number of Ss in the two groups, (f) the mean percentage of CRs made by the HA group of Ss, (g) the mean percentage of responses made by LA Ss, (h) the difference in percentage of CRs given by the two groups, and (i) the significance level of the differences between the groups expressed in terms of *p* (two-tailed test).

As may be seen, the six Iowa studies provided nine independent comparisons of HA and LA groups at UCS intensities ranging from very weak (.25 psi) to moderately strong ones (2.0 psi). In every study HA Ss responded with a

higher percentage of CRs over the conditioning period than did the LA Ss and in each the difference was significant at the .05 level or better. Since the conditioning curves of the HA and LA groups tended to diverge, it is apparent that the differences in their performance at the end of training were even larger than those shown in Table 1. Thus in the case of these six studies the differences were approximately 20% larger over the last 20 conditioning trials than for all trials. These larger differences also tended to be slightly more significant. Attention should also be directed to the fact that the numbers of Ss on which the significance values obtained in these studies were based are relatively large, ranging from a minimum of 45 in studies with two groups to a maximum of 120 when four groups were used.

In turning to the non-Iowa studies in Table 2, it will be seen that not a single one of the comparisons involved as many as 50 Ss. The study (No. 3) with the largest number of Ss (36), it is worthy of note, did give results in line with those obtained in the Iowa laboratory. In this study, conducted at Kent State University, HA Ss responded at a significantly higher level (.01) than LA Ss. The

TABLE 2

NON-IOWA STUDIES OF CONDITIONING PERFORMANCE OF HA AND LA SCALE Ss

Non-Iowa experiments	No. trials	Ready signal	UCS (psi)	No. Ss	Percent CR		Difference (H-L)	<i>p</i>
					HA	LA		
1. Hilgard et al. (1951)	60	Yes	1.6	20	35.2	32.4	2.8	—
2. Prokasy-Truax (1959)	20	No	3.0	20	36.0	48.0	-12.0	—
3. Baron-Conner (1960)	80	No	1.6	36	44.9	27.8	17.1	.01
4. King et al. (1961)	80	Yes	1.5	32	37.4	51.6	-14.2	—
5. King et al. (1961)	80	No	1.5	32	66.8	64.0	2.8	—
6. King et al. (1961)	80	Yes	1.14	30	49.0	51.3	-2.3	—

remaining comparisons in this portion of the table, however, failed to support such a finding, that is, there were no significant differences between groups. The first two of the studies probably should not be taken too seriously as they involved only 10 Ss in each extreme group. The variability of conditioning performance among individual Ss is so great that such results would not be unexpected with so few Ss. However, the final three comparisons, all of which were reported by King, Kimble, and their students (1961) present a more formidable array of evidence that is quite contrary to the findings of the Iowa laboratory. Not only did they fail to obtain a significant difference in favor of the HA Ss, but two of the differences, though not significant, were actually in favor of the LA Ss. While only 30-32 Ss were involved in each experiment, the results as a whole are clearly opposed to the Iowa findings.

In addition to the experiments contained in Tables 1 and 2, the findings of three further published studies that bear on the relation between the *MA* scale and performance in simple conditioning should be mentioned. In each of these investigations unselected Ss were run in the experiment and comparisons were subsequently made between Ss who scored in the upper and lower halves of the distribution of scores on the *MA* scale: Two of these studies were primarily interested in the effects of ad-

ministrating shocks on the level of conditioning performance. In the first, Spence, Farber, and Taylor (1954) found that HA Ss ( $N=15$ ) responded at a significantly higher level in Trials 1-40 than LA Ss ( $N=25$ ) under the shock condition ( $p<.01$ ). Under no shock HA Ss also gave more CRs than LA Ss, but with only 10 Ss per group the difference was not significant. In a partial replication of this experiment carried out at Peabody College, Caldwell and Cromwell (1959) compared Ss, 30 from the upper and 30 from the lower halves of the distribution of *MA* scale scores and obtained a significant relation (.05 level) with conditioning performance over Trials 1-40. In each of these experiments, however, conditioning performance over Trials 41-80 was not related either to shock-no shock or to anxiety level. The third and final study that compared above and below average Ss on the *MA* scale was conducted with students from the evening classes of Northwestern University (Spence & Taylor, 1953). Over the 80 conditioning trials, 22 HA Ss averaged 32.2% CRs, 21 LA Ss, 18.2%. The difference was significant at the .04 level.

*Differential Conditioning.* Table 3 presents the data from two reported investigations of the relation between the *MA* scale and level of performance to the positive CS in differential conditioning. As in the studies of Table 1, HA and LA groups represent roughly the upper

TABLE 3

STUDIES OF PERFORMANCE OF HA AND LA SCALE Ss TO POSITIVE CS IN DIFFERENTIAL CONDITIONING

Study	No. trials	Ready signal	No. Ss	Percent CR		Difference (H-L)	<i>p</i>
				HA	LA		
1. Spence-Farber (1953)	31-50	Yes	36	50.0	39.7	10.3	.05
2. Spence-Farber (1954)	31-50	Yes	52	41.1	31.7	9.4	.09
3. Prokasy-Whaley (1962)	26-45	Yes	60	54.5	39.5	15.0	.05
4. Prokasy-Whaley (1962)	26-45	No	70	60.2	56.7	3.5	—

and lower quartiles on the *MA* scale. Spence and Farber (1953, 1954) conducted two separate experiments differing slightly in conditioning methodology. The data presented in the table show the level of response to the positive CS over the last 20 reinforced trials (31-50). As may be seen, in the first experiment, the HA group gave significantly more CRs than the LA group. The second experiment of Spence and Farber, as originally reported, included Ss on the basis of a forced-choice version of the scale or on the basis of the *MA* scale. The data presented here are based only on Ss who were identified as high or low on the *MA* scale. Again, as may be seen from the table, HA Ss responded at a higher level than LA Ss. The difference was not significant at the .05 level by a two-tailed test, but was on the basis of a one-tailed test.

Using a factorial design, Prokasy and Whaley (1962) investigated whether the

relation between performance on the *MA* scale and differential conditioning was a function of the presence or absence of a ready signal in the experimental procedure. Table 3 presents the percentage of CRs given to the positive CS on the last 20 reinforced trials (26-45). As may be seen, when a ready signal was employed HA Ss gave a significantly larger number of CRs than did LA Ss. In the absence of a signal, however, the superiority of the HA group was small and not significant.

#### UNPUBLISHED DATA INVOLVING UNSELECTED SUBJECTS

The third set of studies providing relevant evidence is presented in Table 4. Included in this group are the findings of two previously published studies (No. 1 & 2) that were concerned with investigating among unselected Ss the relation between conditioning measures and physiological indices of emotionality

TABLE 4

IOWA STUDIES INVOLVING UNPUBLISHED DATA FROM POSTSELECTED Ss

Study	No. trials	Ready signal	UCS (psi)	No. Ss	Percent CR		Difference (H-L)	<i>p</i>
					HA	LA		
1. Runquist & Ross (1959)	41-80	+	1.0	27	70.0	44.3	25.7	.05
2. Runquist & Spence (1959)	41-80	+	1.0	34	57.0	38.2	18.8	.05
3. Individual differences: (unpubl.)	1-80	+	1.0	52	41.8	45.5	-3.7	—
4. Extinction: (unpubl.)	1-30	—	2.0	39	64.3	53.6	10.7	.05
5a. Conditioning problem: Male (unpubl.)	31-60	—	.6	67	61.8	47.6	14.2	.05
5b. Conditioning problem: Female (unpubl.)	31-60	—	.6	31	71.0	59.0	12.0	—

(e.g., heart rate changes, GSR, and muscle action potential). As the scores on the *MA* scale were available it was possible to examine the conditioning performance of *Ss* who scored in the upper and lower quartiles of the distribution of scores on the scale. As the table shows, HA *Ss* gave a significantly greater number of CRs than LA *Ss* in both of these studies.

The remaining data presented in Table 4 come from three as yet unpublished studies from the Iowa laboratory. Each is identified by a word or label describing the primary interest of the study. In each of these investigations a fairly large number of *Ss*, the *MA* scale scores of whom were available after completion of the experiment, were conditioned under identical circumstances. Again, the findings reported are for *Ss* who scored in the upper and lower quartiles of the *MA* scale distribution. As may be seen, the three comparisons (No. 4, 5a, & 5b) provided by the last two experiments once again show that HA *Ss* responded at a higher level than LA *Ss*. Two of these differences were significant, while the third (that involving the smallest number of *Ss*) was not. Attention should be directed to the fact that a ready signal was not employed in these latter studies. The relation of the use or nonuse of a ready signal to the findings of these studies will be discussed in a later section.

The final item of note in Table 4 is the negative difference found in the study entitled "Individual Differences" (No. 3). This is the only instance among a total of 17 comparisons of HA and LA *Ss* provided by our Iowa studies in which LA *Ss* gave a larger number of CRs than HA *Ss*. The small difference is not, of course, significant.

#### DISCUSSION

Looking at the findings of the studies included in Tables 1, 2, 3, and 4 as a

whole, a number of characteristics may be noted. First, the proportion of instances in which the conditioning performance of the HA group was higher than that of the LA group is much greater than one would expect by chance. Thus, the results of 21 of 25 comparisons were in this direction. If there actually were no relation between the *MA* scale and conditioning performance, the probability of obtaining such a percentage of differences (84%) in the same direction by chance is less than .01.

Secondly, it is clearly evident that the studies with relatively large numbers of *Ss* tended to provide significant differences in favor of HA *Ss*, whereas those with smaller numbers did not. Thus, in the case of comparisons involving 36 or more *Ss*, 65% (11/17) of the comparisons were significant at the .05 level or better on the basis of a two-tailed test. On the other hand, in studies involving fewer than 36 *Ss* only two of the eight comparisons provided a significant result. Moreover, all of the 13 significant differences were in the direction of higher conditioning performance on the part of the HA *Ss*, whereas none of the four obtained differences in favor of the LA *Ss* were significant.

As was mentioned earlier, intersubject variability in eyelid conditioning is exceedingly great. In our experiments the percentages of CRs given by individuals customarily range all the way from zero to very high values. The standard deviations of these percentage measures vary roughly between 20 and 25%. Under such circumstances it is readily apparent that fairly large numbers of *Ss* must be sampled if an adequate test as to whether the manipulated variable produces a difference is to be made. This is particularly the case if the size of the difference is not large relative to the variance of the measures. Apparently, degree of emotionality, as specified in

terms of extreme scores on the *MA* scale, is a relatively minor factor or variable among all those contributing to the intersubject variance of conditioning performance. In this connection, it is interesting to note that in two studies in which emotionality differences and differences in UCS (puff) strength were both variables, the differences between the conditioning performances of the HA and LA Ss were of about the same order of magnitude as those obtained between relatively weak and strong puff intensities (.25 vs. 1.5 psi and .6 vs. 2.0 psi). In the case of both types of comparison, individual performances ranged from very low to high levels in the groups being compared and the overlap was considerable.

In the light of these considerations, it is not at all surprising to find such divergent findings among studies that employ such small numbers of Ss, nor is it difficult to understand why the Iowa studies, with their relatively larger samples, have tended to be more consistent and to give a higher proportion of significant differences in favor of the HA Ss.

A third characteristic of the data provided by these studies is that the relation between the *MA* scale and conditioning does appear to vary from one laboratory to another. This fact is clearly revealed in Table 5, which presents an analysis of the data in terms of their source. As may be seen, we have

divided the studies into three groups, those carried out in the Iowa laboratory and those from other laboratories. The latter, in turn, have been broken down into two subgroups, primarily on the basis of the strikingly different findings obtained in them.

It is clearly evident from the data of Table 5 that the Iowa studies have most consistently demonstrated a relation between conditioning performance and emotionality (*MA* scale). Thus 16 (94%) of the 17 comparisons gave differences in favor of the HA Ss, with 10 (60%) of these differences being significant at the .05 level or better. Contrasting most sharply with these findings are those obtained in the subgroup of non-Iowa studies from Duke and North Carolina. These studies, carried out or supervised by Kimble and his associates, failed to find a significant difference in three comparisons of the HA and LA Ss and, moreover, found that LA Ss responded at a higher level than HA Ss in two of their three experiments. Falling in between, but closer to the results of the Iowa studies, are the findings from the four other non-Iowa laboratories. Thus, in four of five comparisons, HA Ss gave more CRs than LA groups, with two of these being significant. Recalling that two of the studies in this latter group had only 10 Ss per group, we see that two of the three experiments that had groups of 36 or more obtained significant differences.

TABLE 5  
COMPARISON OF DIFFERENCES BETWEEN HA AND LA Ss' CONDITIONING  
PERFORMANCE OBTAINED IN DIFFERENT LABORATORIES

Laboratory	No. comparisons	Mean No. Ss	Percent HA > LA	Percent significant (.05)
Iowa	17	47.0	94	60
Non-Iowa I <sup>a</sup>	5	41.2	80	40
Non-Iowa II <sup>b</sup>	3	31.3	33	0

<sup>a</sup> Obtained in the university laboratories of Stanford, Wisconsin, Pennsylvania State, and Kent State.

<sup>b</sup> Obtained in the university laboratories of Duke and North Carolina.

In summary, we see that when studies with a reasonably large number of Ss (36 or more) are considered, the experiments conducted in the Iowa laboratory and those from laboratories *other* than Duke and North Carolina tend to be in close agreement with each other in pointing to a genuine difference in the conditioning performance of the HA and LA Ss. The studies of Kimble and his associates, on the other hand, do not.

In an effort to ascertain what factors might be responsible for these contrasting findings a comparison was made of the experimental conditions under which the studies were conducted. For this purpose a more detailed description of the experimental situation, procedures, and experiences of the Ss was obtained through correspondence with the investigators involved.

With respect to the conditions under which the experiments were conducted at Iowa, we were greatly influenced by the writer's theoretical interpretation as to why HA Ss could be expected to perform at a higher level than LA Ss. According to this theory, the differences in conditioning performance of these two groups of Ss reflect differences in their level of generalized drive (D), which, in turn, are assumed to be the result of differences in their level of emotional reactivity to the experimental situation and procedures. Accordingly, a deliberate attempt was made in the Iowa studies to provide conditions in the laboratory that might elicit some degree of emotionality. Thus, the experimenter was instructed to be impersonal and quite formal in greeting S and in giving the necessary instructions. On coming into the experiment, S at first saw an impressive array of electronic recording equipment and was then led into an adjoining room in which was located an isolated, screened cubicle. The latter contained a dental chair (sic) in which

S was seated in a reclining position, while a headband was placed on his head and a plastic piece was fastened to his upper eyelid. After completing the instructions, the illumination in the cubicle was reduced to a low level of semidarkness and S was informed that, if the need arose, he could get in touch with E by means of a microphone placed on a stand within his reach. The door to the cubicle and the door leading to the adjoining room in which E worked were then closed and S was left in isolation.

To say the least, these conditions were unusual and strange for Ss. Furthermore, in order to maximize the likelihood that they would have a tendency to arouse some degree of apprehensiveness, only individuals who had no previous experience as an S in psychological laboratory experiments were used in all but one of our studies in Tables 1, 2, and 3. It has been our experience that students are much more likely to be concerned and apprehensive in the first experiment in which they serve. After one or two experiences they become much less fearful and, all too often, quite bored and blasé. Some experimental evidence, that the amount of such previous experience is a factor in experiments comparing the performances of the HA and LA Ss, has been provided by an experiment reported by Mednick (1957). The latter found that experimentally naive HA Ss showed more stimulus generalization of a response than LA Ss, whereas no such difference was obtained in the case of Ss who had served in from two to three previous psychological experiments.

It is evident from this account that the conditions under which our experiments were conducted were designed to arrange for some degree of situation-aroused anxiety or emotionality. While one version of our theory hypothesized that differences in emotionality of the

HA and LA Ss might be chronic, an alternative possibility was that it was dependent upon some degree of stress being present in the situation. We were primarily interested in testing the theory under conditions that maximized the likelihood of differences in the emotional reactions of the two groups of Ss.

In the light of the above discussion, it is interesting to compare the conditions and procedures employed by Kimble and King with those employed at Iowa. First, with respect to physical conditions, there are a number of differences. The two Duke experiments were conducted in a cubicle within a room in which *E* was located. However, instead of being semidark, the cubicle was rather brightly lighted (23 apparent foot-candles). The experiment at the University of North Carolina was also conducted under bright illumination with *S* and *E* being in the same room on the opposite sides of a partition. In place of a dental chair, Ss in the Duke experiments were seated on a secretarial-type chair with chin in a headrest. At North Carolina, *S* sat in a straight desk chair with head free. In addition to these physical differences, most Ss in the Kimble-King experiments had already been in one or two experiments prior to serving in the conditioning experiment.

Considering the differences in the two laboratories in terms of the use of different degrees of isolation of *S* from *E*, semidarkness versus well-lighted room, dental chair (emotionally conditioned cue) versus neutral-type chair, and experimentally naive versus sophisticated Ss, it would seem quite plausible to suspect that the differences in the findings could, in part at least, be due to the fact that the experimental conditions at Duke and North Carolina were not as emotion arousing as those employed in the Iowa experiments. Further experimentation is needed, of course, to check on the role

of such factors for there are other differences between the two sets of experiments that might have played important roles. One important one that should be investigated is the possibility that cultural backgrounds of southern and northern students may lead to a difference in the manner in which they respond to the different items in the *MA* scale.

Still another variable is the *E-S* interaction. Our *Es* were not acquainted with Ss and we have attempted to instill in them an attitude of being quite impersonal at the beginning of the experiment and to avoid making any particular attempt to put *S* at ease or allay any expressed fears. Undoubtedly *Es* differ greatly in the degree to which they are able to achieve and maintain this objective. It is interesting to note that the person who served as one of two *Es* in the single Iowa study that gave negative results ( $LA > HA$ ) has recently completed a doctoral dissertation in which there again was little difference between HA and LA Ss. Unfortunately, there were only a small number of the HA and LA Ss under each of the different conditions of this experiment and the proportion of male and female Ss was not equated so the results are not very helpful. This is, nevertheless, a potentially important variable and should be investigated further, possibly by deliberately manipulating the behavior of *E*. It has recently been shown experimentally that one can markedly increase the conditioning performance of Ss by emotion producing instructions (Spence & Goldstein, 1961).

A final variable to be considered is the presence or absence of a ready signal. In a recent article Prokasy and Whaley (1962) have proposed that the use or nonuse of a "ready-blink" signal is a possible factor with which emotionality might be interacting in these condition-

ing studies. In support of this notion these investigators found, as is shown in Table 2, that in differential conditioning, their HA Ss gave a significantly larger number of CRs to the positive CS than LA Ss when a ready-blink signal was used, whereas only a small, insignificant difference was obtained in its absence. However, before this study even appeared there was evidence against the notion that the difference in conditioning performance of the HA and LA Ss was a function of the presence or absence of a ready signal. As Table 1 reveals, Baron and Conner (1960) had reported a highly significant difference ( $p < .01$ ), while Spence and Weyant (1960) also obtained a significant  $F$  ( $< .05$  level) for HA and LA Ss in experiments that did not employ a ready signal. Further evidence against this view that a ready signal is a necessary condition for the difference to occur is provided by the findings of recent experiments from our laboratory presented in Table 4. As the last three items in this table show, three comparisons between HA and LA Ss in which there was no ready signal all gave differences in favor of the HA group with the two comparisons that involved more than 35 Ss being significant at the .05 level.

Unfortunately, a serious methodological error is present in the early studies that did not employ a ready signal. As Hartman and Ross (1961) have recently demonstrated the latency criterion employed in the Iowa laboratory (Spence & Ross, 1959; Spence & Taylor, 1951) to identify and eliminate "voluntary" responders from the data sample is not applicable when a ready signal is not employed. The use of this criterion was originally based on the finding in experiments using a ready signal and a CS-UCS interval of 500 milliseconds that the latency distribution of responses judged to be similar in form to voluntary

eyelid closures did not overlap to any great extent with that for CRs, which are very different in form. Thus, it was possible to employ this more convenient latency property as the index of whether a response was a voluntary response or a CR. As employed in these studies all responses with latencies between 300 and 500 milliseconds were counted as CRs, while responses whose latencies fell between 150 milliseconds (tone CS) or 200 milliseconds (light CS) and 300 milliseconds were counted as voluntary responses. The Ss who gave 50% or more responses in the voluntary category were eliminated from the sample data.

The reason for so doing was that the behavior of such Ss appeared to obey quite different laws from those holding for Ss who gave few such voluntary responses. Among the differences importantly related to our present concern are (a) the tendency to respond at an extremely high level (approaching 100%) once the first such anticipatory voluntary response is made and (b) the lack of any relation whatever between the frequency of responses given by such Ss and the level of the *MA* scale score. The effect of both of these tendencies is clearly that of reducing the likelihood of obtaining a significant difference in the conditioning performance of the HA and LA Ss, the first by greatly increasing the variance of the conditioning scores and the second by reducing the difference between groups that contained a number of such Ss.

Unfortunately, investigators who did not use a ready signal took over the latency criterion to identify voluntary responders without checking its appropriateness, with the consequence that all such Ss probably were not eliminated from their samples. The Prokasy and Truax (1959) study is especially suspect in this regard as it involved a very strong puff (3.0 psi), a condition that produces

a high incidence of such voluntary responders. When one considers the small number of Ss (10 per group) in this experiment, the negative difference obtained is not too surprising. The presence of one or two more high responding voluntary Ss in the LA group than in the HA group could easily have produced this result. The findings obtained with the no-signal groups in the later study of Prokasy and Whaley (1962) and one of the Duke studies also may be suspected of being affected by the presence of a number of voluntary responders who were not detected by the inappropriate latency criterion that was employed.

An example of the effects of including the data of voluntary responders in a sample is provided by the recent studies (Table 4) from the Iowa laboratory that did not employ a ready signal. Having confirmed the findings of Hartman and Ross (1961) that latency does not differentiate voluntary form responses from CRs when a ready signal is absent, the identification of voluntary Ss in these studies was made in terms of the form of the eyelid responses. If, instead, the latency criterion had been employed none of the voluntary responders detected by the form criteria would have been identified and thus would have been included in the samples. It is worthy of note that the inclusion of six voluntary responders (three in each group) in the case of the comparison involving the male Ss of our most recent experiment (Table 4, No. 5a) would have reduced the reported difference of 14.2% between HA and LA Ss to 11.0%. Correspondingly, the significance of the difference would have been reduced from the .05 level to one of .16. In other words, the addition of only six voluntary Ss to the sample of 67 would have changed the conclusion from one that the

difference is significant to one that it is not.

In concluding this discussion, attention should especially be directed to the point that the behavior of human Ss in classical eyelid conditioning, while simple in form, is complexly determined. Recent research in this and other laboratories has revealed that higher mental processes (inhibitory and facilitatory sets, etc.) play a much more prominent role in human conditioning than has sometimes been realized. Moreover, our lack of knowledge concerning these factors has precluded our controlling them in any satisfactory manner with the result that the intersubject variance of conditioning performance is extremely large. This state of affairs has unfortunate consequences for the testing of simple theories of behavior which specify only associational and motivational constructs (e.g., *H* and *D*) as determiners of excitatory strength of the conditioned responses. Obviously, the additional uncontrolled variables minimize the role of these basic theoretical factors, tending to hide their effects. Under such circumstances it is necessary to have reasonably large samples so that the effects of these confounding variables are more likely to be equalized in the comparison groups. On the basis of the writer's experience with conditioning data, experiments of the type discussed here should be required to have groups of at least 25 Ss. Unhappily, as we have seen, this has all too frequently not been the case.

#### REFERENCES

- BARON, M. R., & CONNER, J. P. Eyelid conditioned responses with various levels of anxiety. *J. exp. Psychol.*, 1960, 60, 310-313.
- CALDWELL, D. F., & CROMWELL, R. L. Replication report: The relation of manifest anxiety and electric shock to eyelid conditioning. *J. exp. Psychol.*, 1959, 57, 348-349.
- HARTMAN, T. F., & ROSS, L. E. An alternative criterion for the elimination of "voluntary"

- responses in eyelid conditioning. *J. exp. Psychol.*, 1961, 61, 334-338.
- HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, 42, 94-99.
- KING, M. S., KIMBLE, G. A., GORMAN, J., & KING, R. A. Replication report: Two failures to reproduce effects of anxiety on eyelid conditioning. *J. exp. Psychol.*, 1961, 62, 532-533.
- MEDNICK, S. A. Generalization as a function of manifest anxiety and adaptation to psychological experiments. *J. consult. Psychol.*, 1957, 21, 491-494.
- PROKASY, W. F., & TRUAX, C. B. Reflex and conditioned responses as a function of manifest anxiety. *Amer. J. Psychol.*, 1959, 72, 262-264.
- PROKASY, W. F., & WHALEY, F. L. Manifest anxiety scale score and the ready signal in classical conditioning. *J. exp. Psychol.*, 1962, 63, 119-124.
- RUNQUIST, W. N., & ROSS, L. E. The relation between physiological measures of emotionality and performance in eyelid conditioning. *J. exp. Psychol.*, 1959, 57, 329-332.
- RUNQUIST, W. N., & SPENCE, K. W. Performance in eyelid conditioning related to changes in muscular tension and physiological measures of emotionality. *J. exp. Psychol.*, 1959, 58, 417-422.
- SPENCE, K. W., & BEECROFT, R. S. Differential conditioning and level of anxiety. *J. exp. Psychol.*, 1954, 48, 399-403.
- SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, 45, 116-119.
- SPENCE, K. W., & FARBER, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, 47, 127-134.
- SPENCE, K. W., FARBER, I. E., & TAYLOR, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, 48, 404-408.
- SPENCE, K. W., & GOLDSTEIN, H. Eyelid conditioning performance as a function of emotion-producing instructions. *J. exp. Psychol.*, 1961, 62, 291-294.
- SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, 58, 376-381.
- SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of UCS as determinants of amount of eyelid conditioning. *J. exp. Psychol.*, 1951, 42, 183-188.
- SPENCE, K. W., & TAYLOR, J. A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, 45, 265-272.
- SPENCE, K. W., & WEYANT, G. F. Conditioning performance of high- and low-anxious Ss in the absence of a warning signal. *J. exp. Psychol.* 1960, 60, 146-149.
- TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, 41, 81-92.

(Received August 15, 1962)

## PSYCHOPHYSICAL CORRELATES OF INTRATYMPANIC REFLEX ACTION<sup>1</sup>

MICHEL LOEB

*United States Army Medical Research Laboratory, Fort Knox, Kentucky*

A review is presented of research on the intratympanic muscle reflexes. Among topics covered are a discussion of the methods employed for observation of reflex action, discussion of the natural protection against traumatic noise afforded by the reflexes and its limitation, and discussion of various types of stimulation (especially acoustic and cutaneous) which may be used to elicit the reflexes. Also described are possible practical uses of artificial reflex elicitation to provide enhanced protection against noise and for certain kinds of clinical diagnosis. Implications for damage risk criteria and for certain areas of psychological and physiological research are discussed.

In the past several years there has been a considerable resurgence of interest in the mechanism of the intratympanic acoustic reflex (AR). As Ward (1961) and others have pointed out, this is true not only because of the potential practical value which artificial elicitation of the reflex might have in protecting the hearing of individuals and in clinical diagnosis but also because of the realization of other possible implications. Recent research indicates that the intratympanic muscle action should possibly be taken into account in explanations of masking, remote masking, the scaling of loudness, mechanisms of attention, and intersensory effects.

### NATURE OF THE REFLEX

The intratympanic muscle reflex is a reflexive response on the part of two small muscles of the middle ear. One, the tensor tympani, runs from the cartilaginous portion of the Eustachian tube to the manubrium or handle of the malleus (hammer); the other, smaller muscle, the stapedius, lies within the

posterior wall of the middle ear and its tendon inserts on the head of the stapes (stirrup).

The tensor tympani is innervated by a branch of the trigeminal nerve and may receive sympathetic innervation (Byrne, 1938). It has also been reported recently that the tensor is innervated by a branch of the glossopharyngeal nerve (Lawrence, 1962). The stapedius is innervated by a branch of the facial nerve. Seemingly the muscles receive little or no sensory innervation (Blevins, 1963). In response to activating stimuli—usually loud sounds—the stapedius and tensor tympani pull in such a manner as to damp, limit, or reduce the extent of the ossicular vibrations and correspondingly damp, limit, or reduce the acoustic input to the cochlea. The direction of pull of the muscles is at right angles to the axes of their respective ossicles in such a fashion that the direction of action of one is almost directly opposed to that of the other (Wever & Lawrence, 1954). It has long been known that the reflex is bilateral. Acoustic stimulation of either ear elicits the reflex in both ears (Hammerschlag, 1899; Pollak, 1886). While it is apparent that the reaction to acoustic stimulation is bilateral, there is some question as to whether the homolateral

<sup>1</sup> This paper is a modified version of a presentation at the Seminar on Middle Ear Function at Fort Knox, Kentucky, 7 May 1962. I am especially indebted to J. L. Fletcher and to other participants in the Seminar for suggestions regarding the material included.

and contralateral responses are equal in magnitude. Detailed descriptions of tensor tympani and stapedius anatomy and physiology are available elsewhere (Hilding & Fletcher, 1960; Kobrak, 1959; Perlman, 1960; Wever & Lawrence, 1954).

It has been suggested that the noise attenuating properties of the reflex are primarily attributable to action of the stapedius muscle (Galambos & Rupert, 1959; Simmons, 1959). As some observers have emphasized, the action of the muscle is somewhat analogous to that of the iris and blink reflexes of the eye. However, because of the sharp rise time and brief duration of certain acoustic stimuli, in comparison to the latency of the reflex, and the relatively long duration of other acoustic stimuli, in comparison with the fatigue time of the reflex, the protection offered the ear by the intratympanic reflex is not as adequate as that afforded the eye by the iris reflex.

#### MEASUREMENT OF MUSCLE ACTION

A remarkable diversity of techniques has been employed to measure the action of the intratympanic muscles. Large contractions may be seen with the naked eye or with a microscope, and somewhat smaller ones may be viewed by observing objects inserted into the muscle tendon. Hensen (1878) made direct observation of the tensor tympani and demonstrated that the reflex was elicited by acoustic stimulation. With the stapedius, an appreciably smaller muscle, direct observation is more difficult. Kato (1913) made myograms of stapodial movements by tying a thread to the muscle and connecting it to a stylus on a rotating drum. Today myograms are usually made by monitoring the output of mechano-electronic transducers fixed in the muscle tendons (Wersäll, 1958).

The movements of the middle ear

muscles, especially those of the tensor tympani, result in movements of the eardrum. Stapedius contractions produce an eardrum extrusion; tensor tympani contraction, a retraction. The drum has been directly observed (Hammerschlag, 1899; Kato, 1913; Lüscher, 1929; Waar, 1923), but small movements are quite difficult to detect. Köhler (1909) attached a small mirror to the drum and observed changes in the reflected light. While it is possible to detect smaller movements with this technique, there is probably some loading of the membrane. Other related methods include measurement of changes in volume or pressure produced by eardrum movement (Mendelson, 1961; Terkildsen, 1960; Tsukamoto, 1934) and changes in acoustic impedance of the ear (Jepsen, 1951; Klockhoff, 1961; Metz, 1951). The acoustic impedance of the ear may be defined as resistance to motion on the part of aural structure when acoustic energy is applied. It is a function of the stiffness, mass, and resistance (friction) of the structures and may be estimated by certain pressure or acoustic measurements at the ear canal (cf. Zwislocki, 1961). Manometric, volumetric, and impedance change techniques are objective (although some experimenter judgment is required regarding presence or absence and magnitude of response) and fairly sensitive. It is possible that the impedance change method is more sensitive to tensor tympani than to stapedius action (Reger, 1960), and it is not always possible to determine both tensor and stapedius response on the same manometric recording (Mendelson, 1961). While the latencies of the muscle reflex vary, overlap is possible, and the changes produced by the two may cancel. Moreover, neither the manometric nor impedance change procedure gives a direct indication of actual protection afforded by the

reflex—the reduction of sound energy impinging on the oval window attributable to action of the acoustic reflex.

#### MEASURES OF ATTENUATION: REDUCTION OF COCHLEAR MICROPHONIC AND CHANGES IN PSYCHOPHYSICAL INDICES

One index of amount of protection afforded by the acoustic reflex is the reduction in cochlear microphonics produced by reflex activation (Galambos & Rupert, 1959; Simmons, 1959; Wever & Bray, 1942; Wiggers, 1937). A reduction in the ipsilateral or contralateral response to a loud tone may be noted after introduction of an activating stimulus. Estimates of maximum attenuation range from 10 to 20 decibels (Galambos & Rupert, 1959) to 40 decibels (Wiggers, 1937). It is reported that protection is greatest for tones below 2,000 cycles per second (Simmons, 1959).

While the cochlear microphonic technique has yielded the most data regarding amount of protection afforded by reflex action, it has certain limitations. Cochlear potentials are not readily recorded from humans, although in the course of certain surgical procedures they have been recorded (Ruben, Bordley, & Lieberman, 1961). A more serious objection is that, at or above a moderately intense level, the cochlear microphonic no longer responds linearly to sound (Davis, 1960) and may even be distorted (Lawrence, Wolsk, & Schmidt, 1962).

Alternatively, attenuation may be estimated by changes in several psychophysical indices—notably absolute threshold, loudness, and temporary threshold shift. Békésy and Rosenblith (1951) are perhaps the earliest to report the use of psychophysical determination to measure attenuation afforded by the acoustic reflex. They introduced a loud pure tone into one ear and ob-

served the attendant change of loudness in a tone presented to the contralateral ear. The changes observed indicated attenuation values of 5–10 decibels. Shapley (1954) introduced a loud thermal noise in one ear and noted changes in the loudness of a tone in the contralateral ear. (Loudness reduction was measured by requiring a subject to match the intensity of a tone presented during the contralateral stimulation to the intensity of a tone in the absence of such stimulation.) Attenuation values ranging from 6 to 15 decibels were obtained. There was some possibility, however, of direct masking due to bone conduction of the thermal noise from one side of the head to the other. Loeb and Riopelle (1960) employed a method in which observers were asked to track their threshold at 125, 500, or 1,000 cycles per second by a Békésy technique with and without a loud 2,000 cycles per second tone introduced into the contralateral ear. The difference in thresholds in the presence and absence of the contralateral 2,000 cycles per second tone was assumed to be an index of reflex attenuation. With test and activating tones at these frequencies, the attenuation through the skull is such that masking should not be a factor. Threshold shifts were quite small—approximately 3 decibels. Threshold shifts obtained by the conventional method of limits were of the same order of magnitude. In a second part of the same investigation a loudness balancing technique similar to that employed by Shapley (1954) was used, but a 2,200 cycles per second tone, rather than a thermal noise, was employed to activate the reflex. The data indicated that the loudness of intense tones was reduced to a greater degree than that of tones of lesser physical intensity. It was suggested, on the basis of these experiments and the discrepancy between the threshold shift data and

earlier experiments employing cochlear microphonics, that the reflex might act as a limiter rather than an attenuator, so that intense sounds may be attenuated considerably and faint sounds relatively little. Wever (1962) states that this hypothesis is contrary to other observations (not enumerated) and suggests instead a sort of constancy mechanism—a central process by which action of the reflex is taken into account. It is difficult to see, however, how this latter explanation could account for the small threshold shifts obtained. It should be pointed out that the results of loudness balance experiments by Reger (1960) do not support the hypothesis of differential attenuation of faint and intense sounds, while data from somewhat similar experiments by Prather (1961) tend to corroborate the Loeb and Riopelle (1960) findings.

An alternate explanation advanced (Loeb & Riopelle, 1960) was that subjects may be unconsciously—or consciously—inhibiting the reflex during the psychophysical task. Smith (1943) and Reger (1960) have described individuals who have voluntary control of the acoustic reflex. According to Reger (1960), only 1 or 2% of the population is so talented, but it is possible that a much larger percentage may inhibit their reflexes unconsciously when it is advantageous to do so. In any event, it has been demonstrated that if loud high frequency thermal noise is introduced into one ear and the resultant contralateral threshold shift is measured, larger shifts (as much as 15 decibels) are obtained than when the activating sound is a pure tone (Loeb & Fletcher, 1961; Ward, 1961). Fatigability of the reflex is also seemingly less when it is activated and maintained with a noise stimulus than when a tone is employed. Clicks are at least as effective in arousing and maintaining the reflex as noise (Fletcher &

Loeb, 1962). Moreover, the difference is attributable to the changing quality of noise and click stimuli rather than their complexity, for square waves are no more effective than pure tones in producing contralateral threshold shift (Fletcher & Loeb, 1962). The fact that changing stimuli are more effective in producing contralateral threshold shift than steady ones might be interpreted as evidence supporting the notion that subjects may be somehow inhibiting reflex action during psychophysical measurements. Presumably with ever-changing activating stimulus, this should be harder to do. A more parsimonious explanation might be that reflex adaptation is very rapid for pure tone stimuli and less so for changing ones. Wersäll (1958) stated that the intratympanic muscle reflexes decline in strength over time, but that presentation of a tone of different frequency brings the contraction back to full magnitude. Simmons (1959) noted that a frequency modulated tone produces a more sustained reduction of cochlear microphonics in the contralateral ear than a steady one. Whether adaptation of the reflex is sufficiently rapid to account for the observed differences in contralateral threshold shift is debatable.

It has been suggested (Ward, 1961) that the contralateral threshold shift technique may in part measure "central masking"—a change in sensitivity due to pre-empting of neural pathways common to the two ears. An auditory central masking effect has been demonstrated by Ingham (1959) at levels below those usually considered necessary for activation of the middle ear muscle reflexes. His results, however, indicate that central masking is to be expected when the stimulus in the masking ear has its primary energy in the frequency region identical or adjacent to that of the test stimulus in the contra-

lateral ear. Extrapolation from his results indicates that no central masking should be expected with the frequencies and intensities of activating and test tones employed in experiments on the acoustic reflexes cited above. However, it is possible that at high intensities some other form of central masking may be present. Studies of contralateral threshold shift on patients lacking the middle ear acoustic reflexes (e.g., patients with Bell's palsy or stapedectomized patients) may serve to evaluate Ward's (1961) hypothesis.

Another method of estimating reflex attenuation is the temporary threshold shift reduction (TTSR) technique (Fletcher & Riopelle, 1960; Fletcher, 1961a, 1961b; Ward, Selters, & Glorig, 1961). In this technique a train of loud impulsive noises sufficiently intense to produce a reversible loss of auditory acuity is presented to the observer, but a tone loud enough to activate the reflex is presented before each impulsive noise. (The impulsive noises themselves arouse the reflex, but the muscles do not achieve full contraction before the impulsive noises terminate.) Such a procedure has been found to reduce considerably the amount of temporary threshold shift (TTS) produced by the traumatic stimulus. The average reduction was comparable to that achieved by the use of earplugs known to attenuate pure tones by amounts ranging from 15 decibels at low frequencies to 45 decibels at high frequencies. An average figure, however, is not very meaningful. More TTS at low frequencies was noted when subjects wore earplugs and more TTS at high frequencies when observers were protected only by reflex elicitation.

Hilding (1961) employed a closely related technique. He demonstrated that losses in cochlear microphonic responses in cats, following exposure to impulsive noise, may be reduced or eliminated by

presenting a tone loud enough to activate the acoustic reflex before each of the impulsive sounds. While such a technique has obvious advantages, it must be used with caution. Simmons (1960a) has shown that cochlear microphonic changes do not necessarily correlate with changes in neural sensitivity.

There have been some recent attempts to correlate certain physiological and psychophysical measures of reflex action. Loeb and Fletcher (1961) measured and correlated contralateral threshold shift and TTSR for 16 observers. Correlations between these were found to be negligible. Mendelson and Fletcher (1962) attempted to determine relationships between these same psychophysical indices and manometric indices of acoustic reflex action. In this instance some data suggestive of functional relationship between TTSR and the manometric techniques were obtained, but the relationships were not simple. A recent article (Loeb & Fletcher, 1962) presents data which may explain these results. It describes experiments in which estimates of reliability of contralateral threshold shift were determined. It was found that while contralateral threshold shift in response to high frequency noise or clicks occurred quite regularly, the magnitude of the threshold shift varied somewhat, and reliability was surprisingly low (approximately .50-.60 immediately after introduction of the activating stimulus with experienced subjects and lower in magnitude at later intervals or with inexperienced observers). Reliability of TTS was somewhat higher (approximately .70-.80 immediately after termination of the traumatic stimulus with experienced observers, lower at later intervals or with inexperienced observers). The TTSR reliability was not determined, but it might be expected to be somewhat lower than that of TTS due to variability in the reflex

action itself. With reliability of this magnitude it is apparent that nonsignificant correlations between the indices might very well be obtained. While similar information for manometric and acoustic impedance change techniques is not available in the form of reliability coefficients, it is known that there is considerable within-observer variability in reflex action (Mendelson & Fletcher, 1962). It is also quite probable that a certain amount of variance is introduced in reading and interpreting the recordings obtained by manometric and acoustic impedance change methods.

#### LIMITATIONS OF REFLEX PROTECTION

It has already been mentioned that the latency of the acoustic reflex is such that little or no attenuation of very short, intense impulsive sounds (e.g., gunfire) occurs under normal circumstances. Various values have been given for the latency of the reflex. Some of this variance is attributable to the fact that the tensor tympani and stapedius have different latencies and some to the fact that latency may be defined either as the time before contraction begins or the time before it is complete. Kobrak (1959) proposed that the term "latency" be applied to the former interval and the term "contraction time" to the latter, and this convention will be followed here. He states that the contraction time of the stapedius is approximately 20 milliseconds and that of the tensor tympani 200 milliseconds. Wersäll (1958) states that latency varies somewhat with stimulus intensity, averaging approximately 7-10 milliseconds for stapedius and 12-16 milliseconds for tensor tympani. (These values are for cats under superficial phenobarbital anesthesia; rather similar values were obtained with rabbits.) Contraction time (three cats) ranged from 22 to 38 milliseconds for stapedius and from 56

to 122 milliseconds for tensor tympani. All of Wersäll's values were obtained by myographic techniques. Klockhoff (1961) reported somewhat higher values, but, as he employed an acoustic impedance change procedure in recording activity and elicited the reflex by nonacoustic means, the values should not necessarily be comparable. With manometric procedures (Mendelson, 1961) it is possible to distinguish stapedius and tensor tympani action because they produce opposite movements of the eardrums, but if the responses tend to overlap it is apparent that an accurate indication of latency and contraction time would be difficult to obtain. Cochlear microphonic experiments (Galambos & Rupert, 1959) have indicated reflex (presumably stapedius) latencies of the order of 10 milliseconds. (Estimates of contraction time were not given and are not easy to determine from the records presented.) In any event, though estimates of latency and contraction time vary, it would seem that stapedius contraction requires at least 10 milliseconds and tensor tympani contraction a longer period. There is some evidence that the tensor tympani response is part of the startle reflex pattern (Klockhoff, 1961) rather than a true response to acoustic stimulation, and that it plays little role in reducing acoustic input to the cochlea (Galambos & Rupert, 1959; Simmons, 1959). If so, then the effective time of response can be taken as being between 10 and 50 milliseconds. Since many impulsive sounds have rise times less than .1 millisecond and durations ranging from 10 to 75 milliseconds (Doelling & Kryter, 1959), the protection against impulsive noise trauma afforded by the reflex is minimal.

Contraction time for the middle ear muscle reflexes in man is still not definitively determined. In a very recent

investigation by Ward (1962b) employing the TTSR technique it was shown that to produce maximum reduction of TTS the activating stimulus should precede the traumatic stimulus by at least 150 milliseconds. This result implies either that most protection is provided by the tensor tympani or that the stapedius is fully contracted only after 150 milliseconds. Since the first interprotection is at variance with the results of other findings (cited above), it appears probable that the contraction time for the stapedius in man exceeds 150 milliseconds.

A second limitation of the reflex attenuation is its gradual adaptation to a continuously presented sound. The rate of adaptation apparently varies with frequency, being relatively abrupt for a high frequency tone and considerably less so for lower frequency tones. Adaptation apparently begins immediately but may not be complete after 1.75 minutes (Simmons, 1959). If the frequency of the reflex eliciting, acoustic stimulus is continuously changing (Simmons, 1959; Wersäll, 1958) or if the stimulus is constantly changing in amplitude like a thermal noise or series of clicks (Fletcher & Loeb, 1962; Ward, 1961), the adaptation time may be very much lengthened. The practical consequence of the adaptation phenomenon is that reflex action not only is ineffective against short impulsive stimuli but also against relatively sustained stimuli. However, it is probable that some protection is offered under most circumstances, as animals with sectioned stapedius muscles show greater losses after exposure to continuous noise (Galambos & Rupert, 1959).

After termination of the activating stimulus, relaxation begins almost immediately, if the stimulus is at reflex threshold intensity, or shortly afterward, if the stimulus is considerably above

threshold (Weiss, Mundie, Cashin, & Shinabarger, 1962). However, studies employing acoustic impedance change (Metz, 1946), myographic technique (Wersäll, 1958), and cochlear microphonic reduction techniques (Galambos & Rupert, 1959) all indicate that some contraction may be present for a period of as much as 1 second, though the relaxation period is generally shorter.

In addition to the differential adaptation rate for high and low frequencies previously mentioned, there is also apparently a differential protection according to frequency (Simmons, 1959; Wever & Bray, 1937; Wever & Lawrence, 1954). In general, it is not entirely clear whether the difference is due to differential elicitation of the reflex by tones of different frequency or differential attenuation of tones of different frequency, though the Wever and Bray experiments suggest that the latter explanation is correct. (They found that when the tendons were pulled by external weights, attenuation was greater for lower frequency tones.) In an experiment by Wever and Vernon (1956) tension along the line of the tensor tympani was related to changes in cochlear microphonic. While some animals showed fairly good relationships between tension and microphonic reduction, others did not. The stapedius activity was apparently not measured to any considerable degree. In any event, present data indicate that little protection may be expected from the reflex at high frequencies, especially above 7,000 cycles per second (Simmons, 1959). Whether this is true if the reflex is artificially excited remains an open question.

The threshold for elicitation of the reflex is usually taken to be 70–90 decibels (Wersäll, 1958). Apparently the threshold is somewhat lower at higher frequencies (Simmons, 1960b; Wersäll,

1958), and it also varies with duration (Weiss, 1962). With repeated stimulation, posttetanic potentiation may occur and the reflex threshold may be lowered approximately 24 decibels (Simmons, 1960b). It should be reiterated that at levels below the reflex threshold the muscles apparently maintain a state of contraction such that some sound attenuation is provided (Lawrence, 1960; Simmons, 1959).

It is usually assumed that the acoustic reflex is completely consensual (Wever & Lawrence, 1954). Studies of nonacoustic elicitation of the reflex indicate that reflexes so elicited may be completely homolateral (Klockhoff, 1961) or partially so (Pichler & Bornschein, 1957). There is also evidence that the reflex to acoustic stimulation may not be as great in the contralateral as in the homolateral ear. An experiment employing the acoustic impedance change technique (Möller, 1961) and another employing a loudness balance technique (Reger, 1960) yielded data best interpreted in this fashion.

#### NONACOUSTIC ELICITATION

It has long been known that contraction of the middle ear muscles may be elicited by nonacoustic means. Wever and Lawrence (1954) reviewed the claims of several investigators who describe voluntary contractions of the middle ear muscles. They suggested (citing Schrapinger) that some (not all) individuals who believe that they are contracting the tensor tympani are actually contracting the veli palatini, a muscle of the palate attached to the Eustachian tube. Smith (1943) reported an individual with voluntary control of middle ear muscles and measured changes of threshold during contraction. It is interesting that the changes in loudness reported during contraction were not necessarily correlated

with threshold changes. Reger (1960) also reported individuals with voluntary control and described results somewhat similar to Smith's. Reger's findings previously cited indicate that only 1-2% of the population has voluntary control of muscle movement.

Since, through conditioning, a number of responses have been shown to be capable of elicitation by previously neutral stimuli, it seems at least possible that the intratympanic muscle response may also be so elicited. Simmons, Galambos, and Rupert (1959) have in fact reported that if light flashes are repeatedly paired with loud auditory stimuli the light flashes alone will eventually elicit the muscle response. Ward and Fleer (1961) on the other hand, found that there was no difference in TTS produced by trains of high intensity clicks with periodic and aperiodic interclick intervals. This indicates that the reflex may not be temporally conditioned.

A number of researchers have reported eliciting middle ear muscle reflexes with cutaneous stimuli. Kato (1913) reported contractions in response to various types of aural and facial stimulation. Wersäll (1959) elicited muscle contraction by mechanical stimulation of the pinna and auditory canal. Klockhoff (1961) reported that electrical stimulation of the auditory canal produced a unilateral stapedius reflex, while tensor tympani action could be elicited only by an orbital air jet. The latter response was interpreted as being a component of a startle reaction. Pichler and Bornschein (1957) report a 10-15 decibel increase in the homolateral threshold, attributed to intratympanic muscle action, as a result of electrical stimulation of the external auditory canal. A small contralateral threshold shift (up to 5 decibels was also reported). Attenuation of lower frequencies was appreciably greater than that

for higher frequencies, and by anesthetic techniques the possibility of direct stimulation of the muscles at the intensities employed was ruled out. Djupesland (1962) elicited homolateral and contralateral stapedius reflexes (indicated by impedance change) and an occasional small tensor tympani response by aural air puffs. It is not certain whether the reflex was due to cutaneous or acoustic stimulation.

Hugelin, Dumont, and Paillas (1960) report data indicating elicitation of response of the intratympanic muscles by action of the mesencephalic reticular system. This is discussed below.

#### APPLICATIONS

The most obvious application of the research reported is in construction of a new kind of ear protective device. Fletcher and Riopelle have applied for a patent on an apparatus which will present tones or other reflex eliciting stimuli a short period before the presentation of impulsive sounds. Such a device would have advantages over other protective devices—for example, earmuffs or earplugs—in that no active cooperation on the part of the subject is necessary. (Anyone who has tried to elicit such cooperation will appreciate the importance of this factor.) It is probable that more effective reflex eliciting stimuli will be discovered. Chisman and Simon (1961) suggest that reflex eliciting tones of gradual onset might be less injurious. However, Fletcher's data (1961a) indicate that his activating tones with abrupt onset did not induce appreciable auditory fatigue and so presumably are not injurious. In addition, there is evidence that tones of abrupt onset elicit the reflex more effectively (Weiss, 1962).

Present indications are that the artificial elicitation of the reflex will be of practical value primarily for protection against impulsive rather than continu-

ous traumatic noise. An attempt to reduce TTS following exposure to loud continuous noise by presenting a moderately loud monaural tone during exposure to the noise was unsuccessful (Fletcher, 1961b). However, when either narrow band noise or a train of clicks was presented monaurally during a 25-minute exposure to broad band noise in the contralateral ear, the TTS produced by the broad band noise was significantly reduced (Fletcher, 1961a). Whether protection of this kind would be noted following a longer exposure and whether a system incorporating the reflex eliciting principle would be practical for long continuous-noise exposures is not known, but further exploration of these problems is planned.

A second major application of the findings is the determination of a criterion for exposure to impulse noise. Rather widely accepted criteria for safe exposure to continuous noise already exist (Glorig, Ward, & Nixon, 1961), but comparable standards for exposure to impulsive noise are not available. In the past, it was generally assumed that as impulsive noises are more closely spaced, opportunity for recovery declines, and for the same number of impulses, the danger to hearing increases. However, if the impulses are spaced together closely enough, the reflex activated by one impulse should reduce the inner ear input resulting from the successive one (Ward, 1962a). Carter and Ball (1962) have actually demonstrated such an effect. They find that intense impulses presented at 5 and 10 pulses per second are less damaging than the same number of impulses presented once per second. Similar results have been reported by investigators at the Army Human Engineering Laboratory at Aberdeen, Maryland.<sup>2</sup>

<sup>2</sup> R. Donley and G. Smith, personal communication, 1961.

A third major application would be the determination of reflex excitability for use in clinical diagnosis. Klockhoff (1961) presented data indicating that if a stapedius reflex may be elicited, either through acoustical or electrical stimulation, some middle ear conduction must be present. Djupesland (1962) presented data indicating that this principle may not be entirely general, but his method of reflex elicitation (aural air puff) differed from Klockhoff's somewhat. Absence of such a response would indicate conductive loss. If a tensor tympani response may be elicited (by orbital air puff stimulation) and the stapedius reflex may not, this is indicative of otosclerosis involving stapes ankylosis. Failure to elicit either reflex indicates a more extensive conductive loss. Failure to elicit the stapedius reflex when auditory function appears normal and the tensor tympani response is readily elicitable may be taken as a diagnostic sign for facial nerve lesion (Jepsen, 1955; Klockhoff, 1961).

#### IMPLICATIONS FOR RESEARCH

It has been suggested that changes in remote masking (masking of low frequency sounds by high frequency sounds) may be attributable in part to acoustic reflex activation (Burgeat & Hirsh, 1961; Ward, 1961). It is probable also that a great deal of the effect usually attributed to direct masking (masking of high frequencies by low) is in part due to acoustic reflex action rather than simply to overlap of activated neurons. If so, the degree of masking should change as a function of time, at least for steady masking sounds. Moreover, our perception of the relative loudness of sounds is in all likelihood influenced by acoustic reflex action (Reger, 1960, Ward, 1961). Reger suggests that the slope of the equal loudness contour for intense pure tones re-

flects this function. Loudness of pure tones might be expected to increase as a function of time due to reflex adaptation, but effects of this kind would be opposed by effects due to adaptation and fatigue.

A number of researchers (e.g., Galambos, Sheatz, & Vernier, 1956; Hernández-Péon, Scherrer, & Juvet, 1956) have demonstrated reduction in evoked potentials elicited by auditory stimuli at various central loci as a function of repeated nonreinforced presentation or exposure to distracting stimuli as well as facilitation of such potentials following pairing with reinforcing stimuli. These phenomena have been explained on the basis of activity of the descending reticular activating system (Hernández-Péon, Scherrer, & Juvet, 1956).

Hugelin, Dumont, and Paillas (1960) report that reticular stimulation reduces dorsal cochlear nucleus potentials in "encéphale isolé" cats. However, if the cats are curarized (thereby eliminating intratympanic muscle action) the effect is not obtained. They conclude, therefore, that the physiological mechanism discussed involves reticular system control of the intratympanic muscles. On the other hand, Mousheigan, Rupert, Marsh, & Galambos (1961) have shown that the habituation, conditioning, and distraction effects described may be recorded in cats without middle ear muscles. The conflict of results may be more apparent than real. Desmedt (1960) has presented data indicating that descending pathways not in the reticular system may be responsible for centrifugal control of auditory input. If this is the case, then even if the reticular system does influence the action of the middle ear muscles, such action may not account in large measure for the peripheral filtering of sensory input reported by various investigators. However, the possibility that the intratympanic mus-

cles may exert some such regulatory influence may not be ruled out at this time.

It follows, further, that possible acoustic reflex action should be taken into account in other contexts. If sensory interaction experiments involving cutaneous facial stimulation and acoustical stimulation are performed, possible elicitation of the reflex and resultant attenuation of the sound should be considered. If a study of backward conditioning employs an air puff or an electric shock presented before a tone, it should be realized that the observer may, because of reflex action, not be experiencing the tone at full intensity. Wherever an intense stimulus, especially a cutaneous stimulus to the face or ear, is applied concomitantly with or shortly preceding an auditory stimulus, the possible effects of the reflex should be taken into account in interpreting findings.

#### REFERENCES

- BÉKÉSY, G. v., & ROSENBLITH, W. A. The mechanical properties of the ear. In S. S. Stevens (Ed.), *Handbook of psychology*. New York: Wiley, 1951. Pp. 1075-1115.
- BLEVINS, C. E. Fine innervation patterns of the middle ear muscles. In J. L. Fletcher (Ed.), *Middle ear function seminar. USA MRL Rep.*, 1963, No. 576, 1-15.
- BURGEAT, M., & HIRSH, I. J. Changes in masking with time. *J. Acoust. Soc. Amer.*, 1961, 33, 963-965.
- BYRNE, I. G. *Studies of physiology of the ear*. London: Lewis, 1938.
- CARTER, N., & BALL, J. Studies of temporary threshold shift caused by high intensity acoustic transients. *Bolt Beranek Newman Inc. tech. Rep.*, 1962, No. 916.
- CHISMAN, J. A., & SIMON, J. R. Protection against impulse-type industrial noise by utilizing the acoustic reflex. *J. appl. Psychol.*, 1961, 45, 402-407.
- DAVIS, H. Mechanism of excitation of auditory nerve impulses. In G. L. Rasmussen & W. F. Windle (Eds.), *Neural mechanisms of the auditory and vestibular systems*. Springfield, Ill.: Charles C Thomas, 1960. Pp. 21-39.
- DESMEDT, J. E. Neurophysiological mechanisms controlling acoustic input. In G. L. Rasmussen & W. F. Windle (Eds.), *Neural mechanisms of the auditory and vestibular systems*. Springfield, Ill.: Charles C Thomas, 1960. Pp. 152-164.
- DJUPESLAND, G. Intra-aural muscular reflex elicited by air current stimulation of the external ear. *Acta oto-laryngol., Stockholm*, 1962, 54, 143-153.
- DOELLING, N., & KRYTER, K. D. Characteristics of noise produced by several contemporary Army weapons. *Bolt Beranek Newman Inc. tech. Rep.*, 1959, No. 630.
- FLETCHER, J. L. Prolongation of the action of the acoustic reflex. *J. aud. Res.* 1961, 1, 306-310. (a)
- FLETCHER, J. L. TTS following prolonged exposure to acoustic reflex eliciting stimuli. *J. aud. Res.*, 1961, 1, 242-246. (b)
- FLETCHER, J. L., & LOEB, M. The influence of different acoustical stimuli on the threshold of the contralateral ear: A possible index of attenuation by the acoustic reflex. *Acta oto-laryngol., Stockholm*, 1962, 54, 33-37.
- FLETCHER, J. L., & RIOPELLE, A. J. Protective effect of the acoustic reflex for impulsive noise. *J. Acoust. Soc. Amer.*, 1960, 32, 401-404.
- GALAMBOS, R., & RUPERT, A. Action of the middle ear muscles in normal cats. *J. Acoust. Soc. Amer.*, 1959, 31, 349-355.
- GALAMBOS, R., SHEATZ, G., & VERNIER, V. G. Electrophysiological correlates of a conditioned response in cats. *Science*, 1956, 123, 376-377.
- GLORIG, A., WARD, W. D., & NIXON, J. Damage risk criteria and noise induced hearing loss. *Arch. Oto-laryngol.*, 1961, 74, 71-81.
- HAMMERSCHLAG, V. Ueber den Tensorreflex. *Arch. Ohrenheilk.*, 1899, 46, 1-13.
- HENSEN, V. Beobachtungen über die Thatigkeit des Trommelfellspanners bei Hund und Katze. *Arch. Physiol.*, 1878, 2, 312-319.
- HERNÁNDEZ-PÉON, R., SCHERRER, H., & JOUVET, M. Modification of electric activity in cochlear nucleus during "attention" in unanesthetized cats. *Science*, 1956, 123, 331-332.
- HILDING, D. A. The protective value of the stapedius reflex: An experimental study. *Trans. Amer. Acad. Ophthalmol. Otolaryngol.*, 1961, 69, 297-307.
- HILDING, D. A., & FLETCHER, J. L. The protective value of the middle ear reflex. *Int. Rec. Med.*, 1960, 173, 369-374.
- HUGELIN, A., DUMONT, S., & PAILLAS, N. Tympanic muscles and control of auditory input during arousal. *Science*, 1960, 131, 1371-1372.
- INGHAM, J. G. Variation in cross masking

- with frequency. *J. exp. Psychol.*, 1959, 58, 199-205.
- JEPSEN, O. The threshold of the reflexes of the intratympanic muscles in normal material examined by means of the impedance method. *Acta oto-laryngol., Stockholm*, 1951, 39, 406-408.
- JEPSEN, O. Studies on the acoustic stapedius reflex in man. Unpublished doctoral dissertation, University of Aarhus, 1955.
- KATO, T. Zur Physiologie der Binnenmuskeln des Ohres. *Pflüg. Arch. ges. Physiol.*, 1913, 150, 569-625.
- KLOCKHOFF, I. Middle ear muscle reflexes in man. *Acta oto-laryngol., Stockholm*, 1961, Suppl. No. 164.
- KOBRACK, H. G. The human ear. Chicago: Univ. Chicago Press, 1959.
- KÖHLER, W. Akustische Untersuchungen. *Z. Psychol.* 1909, 54, 241-289.
- LAWRENCE, M. Acoustic effects of middle ear substitution. *Trans. Amer. Acad. Ophthalmol. Otolaryngol.*, 1960, 64, 235-247.
- LAWRENCE, M. The double innervation of the tensor tympani. *Ann. Otol. Rhinol. Laryngol.*, 1962, 71, 1-14.
- LAWRENCE, M., WOLSK, D., & SCHMIDT, P. Inner-ear response to high-level sound. *J. Acoust. Soc. Amer.*, 1962, 34, 102-107.
- LOEB, M., & FLETCHER, J. L. Contralateral threshold shift and reduction in temporary threshold shift as indices of acoustic reflex action. *J. Acoust. Soc. Amer.*, 1961, 33, 1558-1560.
- LOEB, M., & FLETCHER, J. L. Reliability and temporal course of temporary and contralateral threshold shifts. *J. speech hear. Res.*, 1962, 5, 284-291.
- LOEB, M., & RIOPELLE, A. J. Influence of loud contralateral stimulation on the threshold and perceived loudness of low-frequency tones. *J. Acoust. Soc. Amer.*, 1960, 32, 602-610.
- LÜSCHER, E. Die Funktion des Musculus stapedius beim Menschen. *Z. Hals-Nas. Ohrenheilk.*, 1929, 23, 105-132.
- MENDELSON, E. S. Improved method for studying tympanic reflexes in man. *J. Acoust. Soc. Amer.*, 1961, 33, 146-152.
- MENDELSON, E. S., & FLETCHER, J. L. Some indices of middle ear muscle activity. *USN Air Material Cent. Air Crew Equip. Lab. Rep.*, 1962, No. 473.
- METZ, O. The acoustic impedance measured on normal and pathological ears. *Acta oto-laryngol., Stockholm*, 1946, Suppl. No. 63.
- METZ, O. Studies on the contraction of the tympanic muscles as indicated by changes in the impedance of the ear. *Acta oto-laryngol., Stockholm*, 1951, 39, 397-405.
- MÖLLER, A. Bilateral contraction of the tympanic muscles in man. *Roy. Inst. Technol. Speech Transmission Lab. Rep., Stockholm*, 1961, No. 18.
- MOUSHEIGAN, G., RUPERT, A., MARSH, J. T., & GALAMBOS, R. Evoked cortical potentials in the absence of middle ear muscles. *Science*, 1961, 133, 582-583.
- PERLMAN, H. B. The place of the middle ear muscle reflex in auditory research. *Arch. Oto-laryngol.*, 1960, 72, 201-206.
- PICHLER, H., & BORNSCHEIN, H. Audiometrischer Nachweis nichtakustisch ausgelöster Reflexkontraktionen der Intra-auralmuskulatur. *Acta oto-laryngol., Stockholm*, 1957, 48, 498-503.
- POLLAK, J. Ueber die Function des Musculus tensor tympani. *Med. Jb.*, 1886, 82, 555-582.
- PRATHER, W. F. Shifts in loudness of pure tones associated with contralateral noise stimulation. *J. speech hear. Res.*, 1961, 4, 182-193.
- REGER, S. N. Effects of middle ear muscle action on certain psychophysical measurements. *Ann. Otol. Rhinol. Laryngol.*, 1960, 69, 1-20.
- RUBEN, R. J., BORDLEY, J. E., & LIEBERMAN, A. T. Cochlear potentials in the human being. *J. Acoust. Soc. Amer.*, 1961, 33, 830.
- SHAPLEY, J. L. Reduction in the loudness of a 250-cycle tone in one ear following the introduction of a thermal noise in the opposite ear. *Proc. Ia. Acad. Sci.*, 1954, 61, 417-422.
- SIMMONS, F. B. Middle ear muscle activity at moderate sound levels. *Ann. Otol. Rhinol. Laryngol.*, 1959, 68, 1-18.
- SIMMONS, F. B. Middle ear protection from the acoustic trauma of loud continuous sound: An electrophysiological study in cats. *Ann. Otol. Rhinol. Laryngol.*, 1960, 69, 1-9. (a)
- SIMMONS, F. B. Post-tetanic potentiation in the middle-ear muscle acoustic reflex. *J. Acoust. Soc. Amer.*, 1960, 32, 1589-1591. (b)
- SIMMONS, F. B., GALAMBOS, R., & RUPERT, A. Conditioned response of middle ear muscles. *Amer. J. Physiol.*, 1959, 197, 537-538.
- SMITH, H. D. Audiometric effects of voluntary contraction of the tensor tympani muscle. *Arch. Oto-laryngol.*, 1943, 38, 369-372.
- TERKILDSEN, K. Acoustic reflexes of the human musculus tensor tympani. *Acta oto-laryngol., Stockholm*, 1960, Suppl. No. 158, 230-238.

- TSUKAMATO, H. Zur Physiologie der Binnenohrmuskeln. *Z. Biol.*, 1934, 95, 146-154.
- WAAR, A. C. Mikroskopische Wahrnehmungen der Funktion der Mittelohrmuskeln beim Menschen. *Acta oto-laryngol., Stockholm*, 1923, 5, 335-358.
- WARD, W. D. Studies on the aural reflex: I. Contralateral remote masking as an indicator of reflex activity. *J. Acoust. Soc. Amer.*, 1961, 33, 1034-1045.
- WARD, W. D. Studies on the aural reflex: II. Reduction of temporary threshold shift from intermittent noise by reflex activity; implications for damage-risk criteria. *J. Acoust. Soc. Amer.*, 1962, 34, 234-241. (a)
- WARD, W. D. Studies on the aural reflex: III. Reflex latency as inferred from reduction of temporary threshold shift from impulses. *J. Acoust. Soc. Amer.*, 1962, 34, 1132-1137. (b)
- WARD, W. D., & FLEER, R. E. Temporal conditioning of the auditory reflex. *Amer. Psychologist*, 1961, 16, 441.
- WARD, W. D., SELTERS, W., & GLORIG, A. Exploratory studies on temporary threshold shift from impulses. *J. Acoust. Soc. Amer.*, 1961, 33, 781-793.
- WEISS, H., MUNDIE, J. R., CASHIN, J. L., & SHINABARGER, E. W. The normal human interaural muscle reflex in response to response to sound. *Acta oto-laryngol., Stockholm*, 1962, 55, 505-515.
- WERSÄLL, R. The tympanic muscles and their reflexes. *Acta oto-laryngol., Stockholm*, 1958, Suppl. No. 139.
- WEVER, E. G. Hearing. *Annu. Rev. Psychol.*, 1962, 13, 225-250.
- WEVER, E. G., & BRAY, C. W. The tensor tympani muscle and its relation to sound conduction. *Ann. Otol. Rhinol. Laryngol.*, 1937, 46, 947-961.
- WEVER, E. G., & BRAY, C. W. The stapedius muscle in relation to sound conduction. *J. exp. Psychol.*, 1942, 31, 35-43.
- WEVER, E. G., & LAWRENCE, M. *Physiological acoustics*. Princeton: Princeton Univer. Press, 1954.
- WEVER, E. G., & VERNON, J. A. The control of sound transmission by the middle ear muscles. *Ann. Otol. Rhinol. Laryngol.*, 1956, 65, 5-14.
- WIGGERS, H. C. The functions of the intra-aural muscles. *Amer. J. Physiol.*, 1937, 120, 771-780.
- ZWISLOCKI, J. Acoustic measurement of the middle ear function. *Ann. Otol. Rhinol. Laryngol.*, 1961, 70, 1-8.

(Received September 4, 1962)

## NOTE ON NEED COMPLEMENTARITY IN MARRIAGE<sup>1</sup>

GEORGE LEVINGER  
*Western Reserve University*

This note reviews some conceptual and methodological issues concerning the need-complementarity hypothesis discussed in Tharp's (1963) paper on psychological patterning in marriage. It is suggested that, in order to make a valid assessment of the hypothesis, it is necessary: to revise the accepted distinction between need complementarity and need similarity, to provide a more explicit theoretical basis for deciding which needs can be expected to be complementary, and to distinguish operationally the different sources of need gratification.

Tharp's (1963) recent review of studies on the psychological patterning in marriage presents incisive observations about knowledge in this area. His comments on hypotheses and on findings are extremely useful.

It seems necessary, though, to supplement his discussion of Winch's hypothesis on complementary needs in mate selection. To quote Tharp (1963), "In marriage research, no other hypothesis produced in the last decade has been as influential [p. 107]." Yet, after 3 pages of methodological critique of Winch's study, Tharp concludes "the complementary-need hypothesis as now stated is untenable [p. 110]."

This writer agrees with Tharp's general conclusion. Nevertheless, in order to understand why the evidence in favor of Winch's hypothesis is thus far of dubious quality, it is important to go beyond pointing at methodological weaknesses in existing studies. One must also examine the conceptual issues that are involved.

Three issues to which little previous notice has been drawn are the following

<sup>1</sup> The writer's concern with these issues has arisen in the course of research on marital interaction, supported largely by Grant M-4653 from the National Institute of Mental Health, United States Public Health Service. Discussions with Barbara Allan, who is doing empirical research related to the hypothesis, have been helpful.

ones. 1. There exists a logical confusion in the accepted conceptual distinction between *complementarity* and *similarity* of needs. 2. There exists no explicit theoretical basis for deciding *which needs are complementary with which others*, and existing studies are obscure in their basis of selecting such need pairs. 3. The conceptual distinction between *internal* and *external* sources of need satisfaction has been ignored operationally in most research studies that ostensibly bear on the complementarity hypothesis.<sup>2</sup>

The significance of the need-complementarity hypothesis has been recognized not only by those interested in marriage and family relationships, but also by other social psychologists concerned with interpersonal behavior in general (e.g., Newcomb, 1956; Schutz, 1958; Thibaut & Kelley, 1959). From a theoretical standpoint, it is only reasonable to expect that the more Person A wants to give what Person B wants to receive (and the more B wants to give what A wants to receive), the easier it will be for the pair to form and to maintain a viable relationship. The more that such conditions exist, the greater will be the mutual reward of the pair members and the lower their cost of

<sup>2</sup> Some other important conceptual issues pertaining to need complementarity have been treated by Rosow (1957).

achieving it (Thibaut & Kelley, 1959).

1. The Distinction between Complementarity and Similarity: Complementarity may be defined as “mutually supplying each other’s lack,” or “serving to fill out or complete” (Webster’s, 1946). Many of the most productive human relationships are complementary ones: male and female in the sex act, seller and buyer in the exchange, and so forth. In most examples that one can think of, the parts that “complete” the relationship are different in either quality or quantity. And this has given rise to the belief—either implicitly or explicitly expressed in the marriage literature—that complementarity and similarity are mutually exclusive attributes of interpersonal relationships.

Let us examine Winch’s definition of complementarity. If two persons, A and B, are interacting, the resulting gratifications of both are considered “complementary” if one of the following conditions is satisfied (Winch, Ktsanes, & Ktsanes, 1954):

- Type I: “the need or needs in A which are being gratified are very *different in intensity* from the same needs in B which are also being gratified [p. 243].”
- Type II: “the need or needs in A which are being gratified are *different in kind* from the need or needs being gratified in B [p. 243].”

According to Winch and his colleagues, if A is high on need Dominance and B low on this same need, Type I complementarity would occur. They also assert that Type II complemen-

tarity would arise where A is high on need Dominance and B high on need Deference or on need Abasement.

The above definition presents some sufficient conditions for a complementary relationship, but it omits other such conditions. Thereby, it leads one erroneously to the conclusion that if A’s and B’s needs differ *neither* in kind *nor* in intensity, then the relationship is not complementary.

Once apprehended, the error of this impression is clearly demonstrable. Let us assume the existence of a person, the perfect embodiment of Aristotle’s ethical man, all of whose needs exist in perfect moderation. His need scores, on whatever scale one might care to use, fall exactly at the midpoint of the continuum. If Winch’s definition were adequate, one would have to conclude that none of this man’s needs could be complemented, since needs of *different* intensity would be either too high or too low. Such a conclusion seems untenable.

More tenable is the idea that A’s and B’s needs, *same* in kind and *equal* in intensity, will complement one another properly when both members of the pair possess the need in *moderate* quantity. Table 1 illustrates the point for Type I complementarity of the need for Dominance.

The implications of this reformation are both conceptual and methodological. Regarding the former, it is advisable that this important hypothesis be securely anchored to a sound logical foot-

TABLE 1  
COMPLEMENTARITY, SIMILARITY, AND DIFFERENCE RELATIONS IN A NEED PAIRING

Amount of A's need for Dominance	Amount of B's need for Dominance		
	High	Moderate	Low
High	Similar	Different	Complementary
Moderate	Different	Complementary and similar	Different
Low	Complementary	Different	Similar

ing. Until that is done, fruitful research on this subject would be rather difficult to recognize.

Concerning the latter, one may briefly observe that the so-called negative findings in studies of the need-complementarity hypothesis (e.g., Bowerman & Day, 1956; Katz, Glucksberg, & Krauss, 1960; Schellenberg & Bee, 1960) cannot be considered negative merely on the basis of low correlations between partners' needs of the same kind (ignoring for the moment the relation between needs of a different kind). For any given need, if *in general* both partners should indicate it in moderate intensity, then the correlation between the partners' amounts would be reduced artificially to zero, and yet there would be complementarity indeed.

2. Instances of Hypothesized Complementarity: Whereas the first issue is a relatively simple matter of logic, the second is a more complex conceptual and empirical matter. It concerns the question of where the domain of complementarity exists. For what needs can either Type I or Type II complementarity be postulated?

It seems sensible to say that need for Dominance is a case of Type I complementarity; that is, that two partners are better off when one of them needs to make influence attempts and the other to accept them. However, it is much harder to understand the grounds on which Winch et al. (1954) hypothesize Type I complementarity for other needs, such as need Achievement; it is obscure why A and B would be better off when one is high and the other low on need Achievement, as opposed to both being either high or low. For a maximal coordination of their efforts, maximum similarity of achievement motivation might well be most advisable.

It is also obscure how Winch et al. (1954) reason about Type II comple-

mentarity. For example, they hypothesize that Abasement-Hostility and Abasement-Nurturance are both complementary need pairs (p. 246). Winch's published work does not provide a clear theoretical rationale for enabling one to decide what kind of complementarity might be asserted for what kind of need under what conditions.

A more explicit approach toward this problem occurs in Schutz' (1958) more limited theory of need compatibility. Schutz confines his concern to three dimensions of need—inclusion, control, affection—all specifically *interpersonal* needs. While confining himself purely to instances of Type I need complementarity, Schutz distinguishes carefully between A's wishing to *express* behavior and wishing to *receive* behavior along any given dimension. Concerning need for Control, for example, A's and B's needs are defined as compatible to the degree that (a) A wishes to *express* the same amount of control toward B that B wishes to *receive*, and also (b) to the degree that A wishes to *receive* the amount that B wishes to *express*.

Schutz not only presents a useful definition of several kinds of compatibility, but he also has offered some formulae for deriving meaningful scores on need compatibility, together with operations indicated by FIRO-B (Schutz, 1958).<sup>3</sup> So far, only one study of mate selection has reported the use of Schutz' instruments (Kerckhoff & Davis, 1962), but its results lend some support to a hypothesis of need complementarity. It was found that, for pairs of college students who had dated each other longer

<sup>3</sup> One difficulty in Schutz' approach is that his test refers only to the respondent's general peer relations. As presently constituted, his measure is not appropriate for use with married couples; and is more difficult to translate adequately than, say, the items of the Edwards Personal Preference Schedule (Edwards, 1953).

than 18 months, couples with high need complementarity on Schutz' dimensions made more "progress" in their courtship than couples low in such complementarity.

3. Internal versus External Sources of Need Gratification: The third issue refers to the source of satisfaction of any person's needs. An implicit assumption in marriage studies is that need-fit pertains primarily to the marital relationship itself. Yet marriage involves more than a personal relationship between two individuals. The partners are simultaneously involved, both individually and as a couple, in interactional situations in the external environment. There is a substitutability of sources of need gratification for each person's needs. Proverbially, it is accepted that the man whose dominance needs are frustrated at the office comes home to release his anger on wife, child, or dog. Or, conversely, the henpecked husband at home goes out to work to boss his employees. It is by no means necessary that individuals use the same pattern of need gratification within and outside the marital relationship.

Although Winch et al. (1954) mentioned the distinction between need gratification "within the marriage" versus that "in all other situations," this distinction is unclear in their own published data analyses (e.g., Winch, 1958). Furthermore, in the work of other researchers (Bowerman & Day, 1956; Blazer, 1963; Katz et al., 1960; Schellenberg & Bee, 1960), the distinction is almost entirely ignored. The latter studies, in employing the Edwards Personal Preference Schedule (Edwards, 1953), used a need measure specifically designed to measure needs in the general environment of the college student's peers. Their negative findings merely demonstrate, therefore, that marriage partners do not show any consistent cor-

relation between such *general* needs. These studies can hardly claim to test the complementarity of these needs in the specific *marriage* situation.

Parenthetically, it is interesting to note that when Katz et al. (1960) rewrote the Edwards Personal Preference Schedule items describing *general* need for Nurturance and need for Succorance to pertain *specifically* to spouse relations, they obtained their most significant findings for complementarity. Until investigators take account of this issue more directly in their measurement of specific needs, there will be no further evidence to support the original hypothesis by Winch and his colleagues.

This note has attempted to illuminate certain additional issues concerning the need-complementarity hypothesis, omitted in Tharp's (1963) review of marriage research. In this writer's opinion, the hypothesis, when properly conceived and operationalized, does have a viability that gives promise for continued study.<sup>4</sup>

<sup>4</sup> While this paper was in press, additional evidence concerning the effects of *one kind* of need complementarity on marital interaction has been published (Katz, Cohen, & Castiglione, 1963).

## REFERENCES

- BLAZER, J. A. Complementary needs and marital happiness. *Marriage fam. Liv.*, 1963, 25, 89-95.
- BOWERMAN, C. E., & DAY, BARBARA R. A test of the theory of complementary needs as applied to couples during courtship. *Amer. sociol. Rev.*, 1956, 21, 602-605.
- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1953.
- KATZ, I., COHEN, M., & CASTIGLIONE, L. Effect of one type of need complementarity on marriage partners' conformity to one another's judgments. *J. abnorm. soc. Psychol.*, 1963, 67, 8-14.
- KATZ, I., GLUCKSBERG, S., & KRAUSS, R. Need satisfaction and Edwards PPS scores in married couples. *J. consult. Psychol.*, 1960, 24, 205-208.

- KERCKHOFF, A. C., & DAVIS, K. E. Value consensus and need complementarity in mate selection. *Amer. sociol. Rev.*, 1962, 27, 295-303.
- NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, 11, 575-586.
- ROSOW, I. Issues in the concept of need-complementarity. *Sociometry*, 1957, 20, 216-233.
- SCHELLENBERG, J. A., & BEE, L. S. A re-examination of the theory of complementary needs in mate-selection. *Marriage fam. Liv.*, 1960, 22, 227-232.
- SCHUTZ, W. C. *FIRO: A three-dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.
- THARP, R. G. Psychological patterning in marriage. *Psychol. Bull.*, 1963, 60, 97-117.
- THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups*. New York: Wiley, 1959.
- Webster's collegiate dictionary*. (5th ed.) Springfield, Mass.: G. & C. Merriam, 1946.
- WINCH, R. F. *Mate-selection: A study of complementary needs*. New York: Harper, 1958.
- WINCH, R. F., KTSANES, T., & KTSANES, VIRGINIA. The theory of complementary needs in mate-selection: An analytic and descriptive study. *Amer. sociol. Rev.*, 1954, 19, 241-249.

(Received May 8, 1963)

## REPLY TO LEVINGER'S NOTE

ROLAND G. THARP

University of Arizona

Levinger's extension of the complementary-needs (CN) hypothesis is assessed. Although superior to previous CN formulations, it remains a vague statement of faith, lacking theoretical precision and empirical support. The study of needs—complementary or otherwise—is not likely to be fruitful in marriage research. Examples from group dynamics, statistical learning theory, operant theory, personality theory, and phenomenology all demonstrate a general movement in psychology away from the examination of stable traits and motives and toward the use of stimulus situations as predictive constructs. Husbands and wives are spurred to action by stimuli which are not operative in their lives as bartenders and secretaries. Marriage research must specify the stimuli which are specific to spousehood.

Levinger's "Note on Need Complementarity in Marriage" has performed the valuable service of clarifying the conceptual basis on which the social-psychological construct of need complementarity rests. His own reformulation takes account of more permutations than does Winch's system, and he has demonstrated that those who would pair people on the basis of *needs* must take account of complementarity, similarity, difference, compatibility, etc. In addition, he points out that no theoretical basis exists whereby needs may be paired, and we are agreed that attempted pairings have failed to predict social-psychological events.

Thus Levinger has brandished the knife of conceptual clarity, and in the process, committed euthanasia. For he has enabled us to see that a complementary-need hypothesis can only state, in effect, that *some* combinations of *some* needs in *some* pattern of quantity and quality will eventuate in something satisfactory. And since there is as yet no convincing empirical support for *any* such combination, the CN hypothesis remains but a vague faith.

This faith is, I believe, based on the common-sense assumption that some sorts of people combine into more lasting, productive, and mutually satisfying

groupings that would other people. Of course. That it should be otherwise is inconceivable. But should we assume that this matching is on the basis of motives? Levinger argues that the CN hypothesis is viable when needs *specific to the marriage situation* are under investigation. Far from omitting this issue, my original paper (Tharp, 1963b) states that "assessment of needs not specific to marriage is clearly not the logical entree to predictive study [p. 108]." Levinger and I are agreed that the constructs under scrutiny must be *role specific*.

The methodological ramifications of this position are as follows. The concept "role" pertains to expectations, perceptions, and actions specific to a particular situation. Thus it is inappropriate, in social-psychological research, to continue the very operations traditionally employed to assess needs. To abstract response tendencies to a wide variety of stimuli (whether questionnaire items or ambiguous pictures) necessarily brings a loss of the precision required to predict to a specific role.

But the more serious challenge to the CN hypothesis should be made on the basis of the constructs which it attempts to order. We must take account of specific stimulus situations, rather than

attempt to predict from "stable" intra-organismic states. This is a development apparent in the entire enterprise of behavior analysis. Let us consider several examples, the first from the area of social psychology, in which examination of stable traits has given way to a closer inspection of specific conditions (Cartwright & Zander, 1960):

The conception of leaders as people who possess certain distinctive traits has not proved to be satisfactory. A "new view" of leadership is emerging which stresses the performance of needed functions and adaptability to changing situations. According to this conception, groups are (or should be) flexible in assigning leadership functions to various members as conditions change [p. 492].

As another example, in learning theory, drive state (motive, need) was once held accountable as a principal multiplicative constituent of predictive equations; or as the "organizer" of guiding internal stimulation sequences. Current statistical learning theory considers drive inducing operations to affect only the make-up of the stimulus set, the rate of sampling of stimulus set elements, or the probability that any one element will be sampled on any single trial (Estes, 1959).

The most radical statement of this movement toward stimuli is found, of course, in operant theory, where all internal states are extirpated from explanatory statements. Because of the organism's reinforcement history, the appearance of a discriminative stimulus results in a response. Thus the stimulus itself is motivating.

Other areas of psychological theory have profited from greater emphasis on stimulus situation, however, without abandoning their traditional concern with cognitions, perceptions, and expectations. In personality theory, the classical psychoanalytic position of heavy concentration on instinctual drive as the determiner of the form of indi-

vidual behavior is undergoing steady revision. The neo-Freudians, the ego psychologists, and more recently Schachtel in his treatment of the allocentric mode of perception, have demonstrated that the most satisfying and productive human experiences are not to be found in need-dominated behavior but in the appropriate transactions with fully apprehended objects and situations (Schachtel, 1959).

As yet another example, consider the phenomenological approach to man; phenomenology would appear to be the antithesis of operant theory, yet the two are in basic agreement as to the necessity for greater concentration on stimulus conditions rather than on internal states (van den Berg, 1955):

[The psychologist] never learns to know the *subject* better than by going to the *objects*, to the things of his world . . . the relationship of man to world is so profound, that it is an error to separate them . . . Who wants to describe man should make an analysis of the "landscape" within which he demonstrates, explains and reveals himself [p. 32].

Let us then profit from the common experience of our colleagues. It is entirely conceivable that a man or woman is spurred to action as spouse by stimuli which are not operative in these individuals' lives as bartender or secretary. The first order of business for marriage research is the specification of the stimuli: what *is* it to be a spouse? We must be able to describe and quantify the various expectations, reciprocations, rights, and obligations which comprise the situation of spouseness (Tharp, 1963a). Perhaps some facets of this role structure will turn out to be complementary; it is more likely that some will, and others will not. But before concerning ourselves with patterning of elements, we had best identify the elements themselves.

These issues are no different in marriage research than in the investigation

of any social group. We will be closer to knowing why some people form better groups than other people, when we are sure what the groups do, and when, and how.

## REFERENCES

- CARTWRIGHT, D., & ZANDER, A. *Group dynamics*. (2nd ed.) Evanston, Ill.: Row, Peterson, 1960.
- ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. General systematic formulations, learning, and special processes. New York: McGraw-Hill, 1959. Pp. 380-492.
- SCHACHTEL, E. G. *Metamorphosis*. New York: Basic Books, 1959.
- THARP, R. G. Dimensions of marriage roles. *Marriage fam. Liv.*, 1963, in press. (a)
- THARP, R. G. Psychological patterning in marriage. *Psychol. Bull.*, 1963, 60, 97-115. (b)
- VAN DEN BERG, J. H. *The phenomenological approach to psychiatry*. Springfield, Ill.: Charles C Thomas, 1955.

(Received May 28, 1963)

# Psychological Bulletin

## CONTEMPORARY PROBLEMS IN ADAPTATION-LEVEL THEORY:

### A SYMPOSIUM

WILLIAM BEVAN  
*Kansas State University*

The following three papers were presented in a symposium on adaptation-level (AL) theory held during the 1962 meetings of the American Psychological Association. A fourth by Harry Helson (1964), dealing with current theoretical issues, included material which was discussed in his 1963 address as a recipient of a 1962 APA Distinguished Scientific Contribution Award.

The 1962 symposium on AL theory is the second organized by the undersigned. The first, offered at the 1958 APA meetings, was concerned with demonstrating the wide range of phenomena amenable to conceptualization within the framework of this point of view. It consisted of five papers dealing with intermodal anchor effects, reinforcement, social attitudes and action situations, and judgment within clinical populations. This program was well received, and, in light of the consistent interest in the AL model, an increasing concern for problems readily accommodated by this theoretical approach, and a growing empirical literature related to this point of view, it seemed desirable that such a program be repeated after 4 years.

The modern principle of AL originated in Helson's experiments on color constancy and contrast. In these experiments, he made the important discovery that these two perceptual phenomena are complementary and that it is possible to

make accurate predictions of the judged color of a test patch from information on three aspects of the stimulus situation: the reflectance of the patch, the lightness of its background, and the hue of the ambient illumination. Extended to other psychophysical dimensions, the AL principle, in the general case, was demonstrated to reflect the integration of three classes of variable: the focal stimulus, background, and residual stimulation. When determinants representing these three classes are properly identified and controlled, accurate predictions have been made for an impressive array of phenomena: anchor effects, series effects, the effects of interpolated stimuli, shifts in indifference point as a function of change in the statistical properties of the order of presentation, interdimensional and intermodal effects, the effects of subliminal stimulation, the effect of context upon judgment, contrast and assimilation, psychophysical bowing, the effect of presentation interval upon apparent magnitude, the role of criterial attributes in judgment, the influence of expectancies upon judgment, the effective intensity of reinforcers, and reaction time and vigilance as a function of patterns of input.

The 1962 symposium was planned as a two-by-two design where the two variables under consideration were breadth of theoretical application and

degree of empirical articulation. Therefore, it included two general theoretical papers and two papers dealing with specific experimental issues. The power of the theoretical model was explored within the context of two areas of inquiry: that of traditional psychophysics and sensory psychology, and that of personality research. The papers by Helson

and Parducci represented the first-named area and those of Goldstone and Goldfarb and Block the second.

#### REFERENCE

- HELSON, H. Current trends and issues in adaptation-level theory. *Amer. Psychologist*, 1964, 19, 26-38.

(Received October 5, 1962)

## SEQUENTIAL EFFECTS IN JUDGMENT<sup>1</sup>

ALLEN PARDUCCI

*University of California, Los Angeles*

The absolute judgment of a stimulus is often affected by the particular order of presentation of the preceding stimuli. The scale of judgment is anchored by the 2 extreme stimuli, dampening these sequential effects except when the range is extended or when conditions favor shifts in the remembered values of the extreme stimuli. An experimental situation was developed to create the latter conditions. Successive identifications were required for long, randomized sequences of the same 2 stimuli. The sequential effects obtained in this situation were used to evaluate alternative adaptation-level (AL) models, involving the simplest interpretations of the theory and also special assumptions about anchoring by extreme stimuli. This work illustrates a direction for further development of the AL theory.

Trial-to-trial variations in adaptation level (AL) have been generally neglected by those concerned with context effects in absolute judgment. Although it is clear that there are systematic fluctuations in the scale of judgment, even after considerable exposure to the stimulus series, the observed shifts appear to be relatively minor. One assumes that if these local shifts in judgment could be reliably assessed, they would be found to reflect the same principles which describe the adjustment to the series as a whole.

This assumption is rooted in the folklore of the psychology of judgment. For example, after classifying a set of weights into absolute categories, like "heavy" and "light," even a naive experimental subject may volunteer that the same weight felt heavier when it followed a lighter weight than when it followed a heavier weight. We call this successive contrast, and it fits in well with the AL approach to judgment. Such contrast has usually been demonstrated with comparative judgments, as when

an unjudged weight is interpolated between the standard and comparison weight (e.g., Guilford & Park, 1931). The trouble is that the absolute judgments of our retrospective subject are not likely to show this kind of contrast. Insofar as his judgments actually vary as a function of the preceding stimulus, he is more likely to say "heavy" following a heavier preceding stimulus—as assimilation.

This kind of assimilation is rarely a very dramatic phenomenon; and even where clearly demonstrable (e.g., Garner, 1953; Sherif, Taub, & Hovland, 1958), it can be assimilated into an AL approach to judgment (Parducci & Marshall, 1962). But it is a problem, one which we have been happy to ignore in our efforts to describe the "bigger" phenomena of judgment. Thus, we typically get ourselves a more reliable dependent variable by averaging all the judgments made to each stimulus—averaging across different stimulus sequences, hoping that sequential effects will balance out. This may have been sound research strategy since sequential effects are relatively small and since this averaging has enabled us to map out various features of the stimulus context from which one can predict the overall

<sup>1</sup> Presented at the annual meeting of the American Psychological Association, St. Louis, 5 September 1962. Research performed by Arthur Sandusky, Judith Brown, and Weyman Reams, and supported by a grant (M 5465) from the National Institutes of Health.

scale of judgment (Helson, 1959; Johnson, 1955).

However, we might now show a bit more deference to common sense and to our subject's phenomenology. We already know that there are large trial-to-trial effects. The initial adjustment to the stimulus context often reflects quite marked shifts in AL, and similar adjustment may occur repeatedly throughout the series of presentations. Why not submit these shifts in judgment to experimental analysis? The payoff may be a better understanding of the grosser phenomena which we obtain by averaging the trial-to-trial effects.

#### EFFECT OF EXTREME STIMULI

Recent psychophysical research suggests the special importance of one of these grosser phenomena, the influence of the end values, the stimulus extremes, as determinants of absolute judgment (Parducci, 1963). There can no longer be much doubt that these values require special weighting in the AL equations. For example, an increase in the extreme upper value produces immediate downward shifts in judgment of the other stimuli in the series. The AL moves up toward the physical value of this new stimulus. However, there is often little change in judgment following subsequent withdrawal of the new extreme stimulus from the series. The AL does not very quickly return toward its normal level for the restricted series (Parducci, 1956). This "direction-of-shift effect" appears consistent with the special role of the stimulus extremes. To some extent, the subject uses his categories to subdivide the range of stimuli into proportionate subranges; and even after many trials in which the stimulus extremes have not been presented, they still anchor his scale of judgment.

The relative permanence of this end-anchoring in simple laboratory situations may tend to obscure trial-to-trial changes in AL. It is as though the two

extreme stimuli were constantly present as standards against which each of the successive stimuli are compared. The changes in the judgments of the different stimuli are relatively small because they are always being compared with the same standards.

Detailed analysis of the initial judgments of a series should show large, systematic shifts in judgment following each presentation of stimuli with values falling outside the range of the previous presentations. However, since adjustments occur so quickly to extensions of the range, there would be a sparsity of data for detailed analysis. Furthermore, adjustments occasioned by the introduction of new values might differ markedly from sequential effects occurring late in a well-practiced series.

Systematic shifts in the remembered values, the traditional points of subjective equality (PSEs), for the extreme stimuli would provide one of the possible bases for sequential effects which occur later in the series. Insofar as the presentations of the series stimuli affect the PSEs for the extreme stimuli, just as interpolated and comparison stimuli affect the PSE for the standard in comparative judgment, sequential effects would occur throughout the series. The first experimental problem is to find a situation for which the remembered values of the stimulus extremes would be relatively unstable. An experimental situation meeting this requirement would also encourage the operation of other habits of judgment. The tendency for subjects to use their alternative categories of judgment with fixed relative frequencies should have greater influence when the anchoring of the extreme values is made less effective (Parducci, 1963). Thus, a run of presentations from one portion of the range of stimuli should produce shifts in judgment away from the categories previously applied to that portion of the range.

# AN EXPERIMENTAL SITUATION

The results to be described here were obtained using a number of different stimulus dimensions, but always with the same general experimental situation. In accordance with the presumed analogy to changes in PSE for comparative judgment, the situation used a relatively small range of stimulus values so that even small changes in the remembered values of the stimulus extremes would markedly affect the scale of judgment. The subject is told, correctly, that there are two stimuli, for example, a light weight (100 grams) and a heavy weight (120 grams). The stimuli are presented one at a time over a series of 300 trials. On each trial, the subject must identify which of the two stimuli (i.e., the lighter or the heavier weight) has been presented. He is given no feedback concerning the correctness of his judgments. Subjects under different conditions receive the two weights with different probabilities (e.g., the heavier weight may come on 80% of the trials, or on 50%, or on only 20% of the trials). In our efforts to find a suitable experimental situation, we have tried (besides lifted weights) lights varying in brightness, tones varying either in loudness or in pitch, and, most recently and extensively, two lights which vary in spatial position—the subject having to judge whether, in a darkened room, a small light has flashed in the left or in the right of the two possible positions (a third, fixed light being located far above these two variable lights).

## ALTERNATIVE MODELS

Consider some alternative models for judgment in this situation. The simplest AL model, consistent with Helson's weighted-mean approach, might treat AL on any trial as the mean of all the stimulus presentations up to and including that trial, the probability of a correct judgment being a linear function

of the difference between the value of the presented stimulus and AL on that trial:

$$AL_n = \frac{n-1}{n} AL_{n-1} + \frac{1}{n} S_n \quad [1]$$

$$P(J_n=1) = K(S_n - AL_n) + C \quad [2]$$

where  $AL_n$  is the adaptation level on Trial  $n$ ,  $S_n$  is the stimulus value (e.g., 1 or 0) on Trial  $n$ ,  $P(J_n=1)$  is the probability that the judgment on Trial  $n$  is "heavier" (or "brighter" or "left"), and  $K$  and  $C$  are empirical constants representing the discriminability of the stimuli and the response bias, respectively. Since the influence of  $S_n$  upon  $AL_n$  varies inversely with  $n$ , this model entails that the probability of a correct response should become increasingly independent of features of the stimulus sequence (e.g., the length of homogeneous runs) as the number of trials increases. This has been the general observation for experiments on absolute judgment, the systematic changes in the judgment of each stimulus being small after an initial adjustment to the series. Although this model appears consistent with a fair body of research, our data from the present situation show little difference between the sequential effects found in the first versus the second 150 trials, even though fairly marked trial-to-trial changes appear throughout the series.

Another AL model<sup>2</sup> seems more consistent with certain developments in Helson's thinking (e.g., 1948). This second model also assumes that AL on each trial is a weighted average of AL on the preceding trial and the stimulus occurring on the present trial. But for this model, the weighting is the same for every trial:

$$AL_n = aAL_{n-1} + (1-a)S_n \quad [3]$$

<sup>2</sup> First presented by Seward (1961) for describing the effects of incentive magnitudes, and independently developed for perceptual judgments by N. H. Anderson (unpublished manuscript).

where  $a$  is an empirical constant which remains constant over trials. The law of judgment is the same as for the first model (Equation 2). Thus, AL shifts toward each successive stimulus, moving a constant proportion (independent of trial number) of the difference between where it was and the value of the presented stimulus. Unlike the first model, this one does not necessarily entail changes between first-half and second-half sequential effects (if the rate of change is large). It does, however, entail that AL approximates the mean of the stimuli with long, randomized series—in this sense explaining this AL phenomenon obtained by averaging over the entire series. It also predicts that the probability of a correct response will decrease as a function of the length of a homogeneous run (e.g., a series of presentations of the lighter weight). That is, AL approaches the value of the repeated stimulus, decreasing the difference between AL and that stimulus. So the likelihood of a correct response should decrease. And this is what happens. Subjects are much more likely to be correct on alternation than on repetition trials, the difference being on the order of 75% correct when the stimulus shifts as compared with only 60% correct on trials when the stimulus repeats (for the degree of stimulus variation used with several different physical dimensions studied for the present situation).

Another prediction from this model is that the probability of being correct on the stimulus breaking a homogeneous run (i.e., on alternation trials) will increase as a function of the length of the broken run. The AL should move toward the repeated stimulus so that the difference between AL and the stimulus which finally breaks the run will get progressively larger—increasing the probability of a correct response when the stimulus run is finally broken. This

model fails here. For every dimension studied, the probability of a correct response appears to be independent of the length of the prior run on the alternate stimulus.

A third derivation (from both of these AL models) is that the probability of each response (e.g., "brighter") is independent of the overall probability of presentation of the associated stimulus. This is an interesting prediction since the stimulus values yield correct discrimination on more than two thirds of the trials for a 50% schedule. One might have expected, for example, more "brighter" judgments when the brighter light was presented on 80% of the trials (as compared with conditions for which it was presented on only 50% or 20% of the trials). Although contrary to common sense, this prediction seems central to the general approach taken by AL theory. The initial results were most intriguing, the mean percentage of "brighter" responses, for example, being within one percentage point for all three schedules (56%). However, subsequent work using judgments of two different positions of lights indicates that this prediction does not always hold, the probability of a "left" response shifting directly with the probability of the light flashing in the left position (.53, .50, and .47, for the .80, .50, and .20 schedules, respectively).

A third model seems to handle the data better than either of the two already described. In accordance with the special role of the extreme stimuli, this model assumes that probability of a correct response is a linear function of the difference between the PSEs or remembered values of the two stimuli which are the end points of the stimulus distribution (there being just the two stimuli in the distribution). The model also assumes that the PSE for the presented stimulus is always the same value, the physical value of the presented

stimulus (e.g., 120 grams), and that the PSE for the stimulus not presented is a weighted average of its value on the preceding trial and the value of the presented stimulus:

$$s_n = a s_{n-1} + (1-a) S_n \quad [4]$$

$$P(J_n = 1) = K(S_n - s_n) + C \quad [5]$$

where  $s_n$  is the PSE for the stimulus not presented on Trial  $n$ . This end-point model does better than the AL models since it not only predicts that the probability of a correct response decreases with increase in the length of a homogeneous stimulus run, but also that the probability of a correct response to a stimulus alternation is independent of the length of the prior run. Furthermore, it correctly predicts that the probability of one of the responses (e.g., "left") will sometimes vary for different stimulus probabilities—depending upon the rate of change of PSE.

Some caution must be expressed with respect to the adequacy of this simple end-point model. The declining function which represents the fall in the probability of a correct response on homogeneous stimulus runs flattens out on the second repetition and sometimes even rises—indicating that there may be an alternation bias in the subject's responses. While there is a general tendency (i.e., in other situations) for judgments to be equally distributed over the different response categories, certain guessing habits may be specific to this simple situation.

In spite of these reservations, it is clear that this situation yields marked trial-to-trial effects, systematic for each subject and also across subjects, and that these effects are associated with particular stimulus sequences. We thus have the opportunity to analyze the basic nature of the shifts in frame of reference which occur with each successive stimulus presentation. Why are the sequential effects so much greater here than for

the more complex experimental situations with which absolute judgments are customarily investigated? Perhaps the use of end points which are themselves difficult to discriminate—only about one difference limen apart—may be crucial. However, the sequential effects for everyday, nonlaboratory judgments, with greatly separated end points, sometimes seem to be of much greater magnitude. Of course, such extreme experiences may be widely spaced in everyday life. The fact that we can get these trial-to-trial changes in a single laboratory session suggests that further investigation of sequential effects may provide a basis for significant developments in the AL approach to judgment.

# REFERENCES

- GARNER, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, 46, 373-380.
- GUILFORD, J. P., & PARK, D. G. The effect of interpolated weights upon comparative judgments. *Amer. J. Psychol.*, 1931, 43, 589-599.
- HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959.
- HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.
- JOHNSON, D. M. *The psychology of thought and judgment*. New York: Harper, 1955.
- PARDUCCI, A. Direction of shift in the judgment of single stimuli. *J. exp. Psychol.*, 1956, 51, 169-178.
- PARDUCCI, A. Range-frequency compromise in judgment. *Psychol. Monogr.*, 1963, 77(2, Whole No. 565).
- PARDUCCI, A., & MARSHALL, L. M. Assimilation vs. contrast in the anchoring of perceptual judgments of weight. *J. exp. Psychol.*, 1962, 63, 426-437.
- SEWARD, J. Beyond canalization. In J. G. Peatman & E. L. Hartley (Eds.), *Festschrift for Gardner Murphy*. New York: Harpers, 1961.
- SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, 55, 150-155.

(Received November 27, 1962)

## A CONCEPTUAL FRAMEWORK FOR THE CLINICAL TEST SITUATION<sup>1</sup>

WILLIAM E. BLOCK  
*Yeshiva University*

The need to account for the adaptive behavior of the patient in the clinical test situation is discussed with respect to the relative rather than absolute nature of test responses. Adaptation-level (AL) theory is proposed as a framework of reference for adaptive behavior of the patient in the clinical test situation, as well as for clinical judgment and prediction. The theory of AL possesses the relativity and operational validity required for dealing with complex clinical phenomena. Implications of the AL model for clinical theory, experimentation, and practice are suggested.

The use of projective techniques is one of the main approaches of the clinical psychologist in understanding the patient before him. The clinical test situation is a complex interaction of a patient with a past history, hopes and fears, and with varying motivations, who must adapt to a projective task characterized by more or less ambiguous stimuli (the ambiguity of which varies with patient expectations) administered in an interpersonal context by a skilled examiner whose idiosyncrasies, in turn, may influence the patient. Hence, the patient's responses can only be regarded as relative to the totality of influences operating in this complex situation. The necessity arises, therefore, for a theoretical framework broad enough to encompass the effects of focal, residual, and contextual factors upon test behavior. This article discusses the possible application of the adaptation-level (AL) paradigm to problems of the clinical test situation.

### CLASSICAL VIEW OF THE CLINICAL TEST SITUATION

The patient's ostensible freedom from inhibitions of propriety, self-censorship, and right-wrong set in confronting the ambiguous task represented by projective stimuli has been regarded as among the chief virtues of this type of test compared with, say, the objective personality inventory in revealing deeper aspects of personality, hidden motivations, etc. The response elicited by the projective technique is viewed as a "projection" of personality in a manner analogous to the way the hidden images of a film strip are brought to life by projection on a screen; another common analog views the projective test as an X ray of underlying personality structure. Protocol variables of main interest to the clinician have been those linked to "task performance," such as responses, scores, and content, with little or no concern about another domain of variables centered on "task approach" or "adaptation."

Projective techniques are largely imbued with face validity and the resulting protocol taken as indicative of absolute traits of the patient. Other working assumptions about clinical testing are the monolithic nature of the test situation, that is, its freedom from ex-

<sup>1</sup> This paper was prepared under National Institutes of Health Grant M-6101, and is an extension of a study reported at a symposium on AL theory held at the 1962 meeting of the American Psychological Association in St. Louis. The title of the symposium paper was: "A Preliminary Study of Adaptation-Level Theory as a Framework for Projective Testing" (National Institutes of Health, M-5118A).

traneous influences, the validity of the concept of "projection," the inferences drawn from responses, and the simple addition of various tests to comprise a "battery."

### CRITIQUE OF THE CLINICAL TEST SITUATION

Despite consensus among clinicians of what transpires in the test situation, a number of unresolved questions emerge about these underlying assumptions.

1. Projection is the accepted explanation of the process by which a subject produces a response to the ambiguous test features. Yet projection is descriptive rather than explanatory in the scientific sense. While it is a rich concept, the what, how, and why of projection remain moot questions. Projection also involves a paradox—the notion of an unconscious illogical psyche making a logical cognitive effort to comprehend the projective task.

2. Aside from the problems inherent in obtaining a test protocol, discussed below, the meaning of the protocol poses another problem. Not only may interpretative approaches vary within a given technique, but also from one technique to another. Interpretation may involve formal, content, dimensional, and holistic aspects. Though all projective techniques are thought implicitly to involve projection, articulation of data for purposes of interpretation from various tests becomes an artistic, intuitive enterprise because of their unique metalanguages. Hence, the clinician more often than not utilizes dubious overarching conceptual canopies, procedurally rather than empirically validated (Block, 1962b), to encompass the variegated data of a multitest battery.

3. Moreover, many projective techniques involve a priori assumptions about the meaning or significance of certain responses as well as their interrela-

tionships without operationally anchoring either the basic concepts or rules of interpretation.

4. Until recently, it was assumed in the classical measurement tradition of Galton that a test sampled a subject's response to invariant stimuli in a standard invariant situation administered by an inscrutable examiner. This monolithic view of the clinical test situation essentially holds that a sample at one time is representative of the patient on all other occasions.

This view becomes untenable in the light of increased sophistication about the nature of psychological testing and a growing number of empirical studies (Masling, 1960) showing that the subject's responses are influenced not only by his personal needs, both residual and transient, but also by his appraisal of such situational variables confronting him as the mode of administration, testing milieu, and characteristics of the examiner.

The gist of this evidence is that the subject's response is a relative one tempered by adaptation to the here-now test situation. And though clinicians are becoming more cognizant of background, focal, and contextual effects in testing, there has been little effort made on their part to integrate adaptive behavior into the overall personality description of the patient. The fault does not lie with clinicians but with the inadequacies of test theories at hand. Projective psychology, as a body of theory, can neither take into account the patient's adaptations nor make use of this important facet of behavior.

5. Another questionable assumption is the objectivity of the examiner. It had been previously thought sufficient merely to provide the clinician with a set of rules of interpretation or principles to insure objective scrutiny of protocols. The evidence is that the clini-

cian's personality does enter into his judgments of test data. Yet no way is known to mitigate this bias since there is no clear understanding of how judgments are formed in the first place.

6. The absolutist tradition of testing assumed that the addition of stimuli (or tests) merely increased the range and variety of responses, and that each test was independent of others in the total battery. The fact that interactional effects occur between other components of the clinical triad (test-testee-tester) suggests the likelihood of interactions among the test stimuli vis-a-vis the patient as well.

7. Most research on projective testing is focused on what is measured to the exclusion of the more fundamental question: How does the projective technique qua measuring instrument elicit data? As the author has suggested elsewhere (Block, 1962b), projective techniques are similar to psychometric experiments albeit with unexplicated rationale. As such, they are susceptible to the same class of errors and judgment biases that pervade any measurement effort. Once again, projective psychology is found inadequate to deal with these psychometric phenomena.

The above critique implies a cogent need for a theoretical framework with construct validity to account for the diverse adaptive phenomena that operate in the clinical test situation. The author would now like to state the case for AL theory in meeting this need based upon theoretical considerations and preliminary experimental findings.

#### AL THEORY AS A FRAMEWORK FOR THE CLINICAL TEST SITUATION

The concept of AL, as formulated by Helson (1948, 1959), provides a framework within which the variables of the clinical test situation may be evaluated on a quantifiable, operational basis. De-

finied operationally, AL is the stimulus that evokes a neutral or indifferent response by the subject. Since the structure of the behavioral field is fixed by the position of the neutral point, the frame of reference is completely defined once the value of AL is established. The AL is the pooled effect of stimuli interacting with each other (weighted-log mean). The existence of a neutral region (AL) is important because it acts as a frame of reference for behavior and determines what reactions will be made to the specific stimuli in any behavioral situation. The AL also involves temporal pooling of residual, contextual, and focal stimuli. By virtue of its mathematization, AL theory permits calculation of the relative contribution of these stimulus sources to a given response or judgment.

Thus AL theory deals with the here-now adaptive behavior of the individual taking into account past experiences, present needs, and future anticipations. These dimensions of past-present-future characterize the temporal aspects of normal existence. The theory of AL can deal with those aspects of behavior which convey a sense of stability to the individual by harmonizing inner needs with the outer demands aimed at establishing a stable, yet dynamically responsive, point of reference for behavior.

Being operationally defined, AL theory is a rigorous scientific system unlike the quasi-mystical concepts of projective psychology. Moreover, since the theory deals with adaptive behavior, it satisfies the requirement of psychological relevance too. And of special importance for the esteemed "subjectivity" of the clinical test situation, AL theory allows for phenomenological and existential aspects of behavior.

Starting from simple psychophysical experiments, Helson and others have

systematically shown the wide applicability of AL theory to other psychological domains such as vision, learning, social psychology, traumatic and neurotic behavior. Helson (1959) has postulated the extension of AL theory to psychological testing in this statement:

conceptually responses to ink blots or test items in personality scales are no different from psychophysical judgment in being influenced by situational factors and momentary adjustment level of the organism [p. 605].

Murstein (1959) utilized the AL framework in analyzing stimulus research on the Thematic Apperception Test, though not demonstrating its application experimentally.

Block (1962a) conducted a preliminary study to test the congruence between AL theory and projective testing. Twenty subjects rated the Rorschach inkblots on the evaluative scale of the semantic differential following the usual free response phase. Factor loadings of each card on the evaluative scale were used as their objective stimulus value. The AL corresponding to the neutral point of pooled ratings was computed and then original mean judgments were postdicted on the assumption that such judgments had taken place with respect to a framework of reference determined by the prevailing AL. An  $F$  test of "goodness of fit" between experimental and theoretical values was significant beyond the .01 level of confidence; this finding was cross-validated on another sample of 13 subjects. It was therefore possible to apply the AL paradigm not merely analogically, but by means of general quantitative, functional relations established between stimulus and behavioral variables in a projective testing situation.

#### BEHAVIORAL DIMENSIONS OF AL

A paramount value of the mathematization of functions is the possibility of

manipulating symbolic concepts to reveal new dimensions of understanding and insight. This has been done in some AL formulations.

First, AL may be described as a pooling or interaction of effects. It is the end result of averaging processes in the organism. On the assumption of isomorphism between behavior and its mathematical expression, Helson (1959) infers that AL exemplifies psychological integration, inasmuch as averaging operations mathematically are special cases of integration. This implies that the subject's confrontation with the clinical test situation, as measured in AL terms, is an index of degree of psychological integration at that moment.

As a corollary of this integrative view, the manner in which residual (past experience), focal (present needs), and contextual (milieu) effects interact tells something about the individual's handling of the adaptive task he confronts in the test situation. This may be illustrated by the mathematical definition for AL where there is pooling of existing levels of stimulation (Helson, 1959):

$$AL = AL_r^m \cdot AL_c^n \cdot S_f^e$$

where AL stands for adaptation level,  $S$  denotes the geometric mean of projective stimuli; the subscript  $r$  for residual,  $c$  for contextual,  $f$  for focal; and exponents  $m$ ,  $n$ , and  $e$  are the respective weighting coefficients of the various stimulus sources contributing to AL. Michels and Helson (1949) have suggested the following psychological equivalents of the weighting exponents which we shall reinterpret for the projective test situation:

$m$ —this is the contribution of past experience, and the fact that it must be taken into account in adapting to any new situation or set of facts makes it psychologically the equivalent of resistance in the test situation. Such a meaning would be most useful perhaps in assessing the

subject's openness to therapy or new experiences via test adaptation.

$n$ —this represents the contribution of the surrounding test milieu, be it human, physical, social, etc., in short, the relativity of responses to the prevailing situation. Psychologically, it may best be described as a measure of the subject's distractibility from the projective task at hand.

$e$ —this exponent represents the contribution to AL of the subject's inner needs, their persistence and strength of manifestation in the face of competing residual and contextual influences of the test situation. In a sense it represents self-adaptation, how one confronts inner tensions; its behavioral counterpart is sensitivity to one's own needs in the face of external demands.

Recapitulating, we may then say that adaptive test behavior integrates past experience, present needs, and background influences. The important implication here is that so-called fixed personality traits sought by clinicians are relative to the test situation in which they are called forth. The practice of searching for absolute traits as such is questionable on theoretical and empirical grounds. One may only hope to find or show relative constancy between given stimuli and responses despite the shift of the zero of functioning (AL) under changing conditions (Helson, 1959).

#### AL THEORY AS A MODEL OF CLINICAL JUDGMENT AND PREDICTION

The clinical psychologist appears to engage in two judgmental processes with respect to the data confronting him, whether test, interview, or observations. The first type of judgment involves description, or classification of data; the second, more complex type, is the prediction of future behavior on the basis of present data.

How clinical judgments are formed has been scrutinized carefully by a number of researchers, one group of which is inclined to relegate clinical decision making to the realm of the artistic and

other such realms, implying refractoriness to scientific explanation; while the opposing group views clinical judgments as rational and scientifically verifiable. The latter have increasingly used mathematical models to describe clinical judgments. Hoffman (1960), for example, proposes both linear-additive and configurational-interactive models, while Bakan<sup>2</sup> offers a probability model.

In the present author's opinion, such mathematical models suffer by being paramorphic; that is, not really tied to the underlying processes purportedly represented. Contrasted with these, AL theory offers an isomorphic model, one having construct validity because of its verifiable relationships between stimuli and responses. Moreover the AL model embraces both interactive and probabilistic aspects of clinical judgment.

As to the first type of judgment—classification or description—the general paradigm of AL theory for psychophysical judgments would seem to apply; for in principle, any set of test responses or qualitative data that can be scaled and dichotomized is amenable to representation in AL terms. Focal, contextual, and residual factors will be contributory to the overall judgment for the given patient.

The theory of AL applied to clinical prediction is more complex and may involve what Smedslund (quoted by Helson, 1959) defined as probability-adaptation; that is, the level of probability expectation that is affectively neutral to the person at a given time and which is represented by the weighted average of all levels of probability expectation influencing the given situation. Thus probability adaptation is the pooling of present and past levels of expectancy. Clinical prediction may therefore be regarded as a statement of proba-

<sup>2</sup> In an address to the Graduate Psychology Club at Yeshiva University in 1961.

bility expectation about future behavior on the basis of previous expectations. For the clinician, the past or residual expectation is his "background of clinical experience" or general base-rate expectancies about phenomena similar to the focal issue; present expectancy about a phenomenon is derived from viewing it against this past experience; future expectancy or prediction is then an interaction of these two probabilities according to the weighted-log mean formulation. The author has planned experiments to test this probability-adaptation model of clinical judgment.

#### PERSONAL CHARACTERISTICS IN CLINICAL JUDGMENT

Numerous studies (Masling, 1960) show that the clinician's judgment of test protocols is influenced by his personal and behavioral characteristics. Thus, another presumed absolute feature of the test situation is found to be a relative one. What has been overlooked in these studies, though, is the effect of the stimuli upon the clinician and his subsequent judgment of a subject's protocol. What is suggested here is that the clinician's judgments about the subject are contaminated, as it were, by his own adaptation to test stimuli. In AL terms, the clinician's self-adaptation to projective stimuli acts as a contextual or background influence in judging the subject's adaptation to these same stimuli. Helson's (1948) general formulation for judgments with anchor stimuli would seem applicable here. This view is supported, too, by the study of Rethlingshafer and Hinckley (1954) showing the effects of judges' attributes (age-weight) on judgment of the subject's age-weight according to the anchor effect described above.

An AL model of clinical judgment may open the way toward greater understanding of factors to be weighed in

making judgments, individual differences among clinicians in judging, and effects of training, personality, and idiosyncrasies of the clinician in the interpretation of test data.

#### AL THEORY AND ASSUMPTIONS IN CLINICAL TESTING

The clinical test battery is usually comprised of several techniques to provide different kinds of information about the patient according to the nature of each component test. Two assumptions surround use of the clinical battery: (a) despite different underlying metalanguages each test contributes to the same overall meaning and (b) each test is independent of the others in the battery.

The first assumption about underlying similarities in the face of surface dissimilarities appears unwarranted because: (a) different metalanguages of the tests probably present different meaning structures to which the patient must adapt and (b) combining tests for interpretation within some global theoretical framework—say ego psychology—is gratuitous. These inherent differences of meaning structure and adaptive tasks can be transcended, however, by viewing the response to the battery and its components within the framework of AL theory. This is in keeping with the growing shift from task performance to task approach. One may then trace the subject's progress in adapting from one task to the next within the battery, each test presenting a unique adaptive task, but contributing to his overall adaptation.

The assumption of independence of tests in the battery seems unwarranted because the subject's responses may reflect successive adaptation to different tests rather than an absolute response true at all times for a given test. The test battery may be viewed as a series of stimuli (tests) in which AL prevails

as a function of the pooling or interaction of the stimuli in the series. Hutt (1947) found variations in Binet IQ depending upon whether subtests were presented consecutively or adapted to the subject's emotional needs of the moment. Heim (1955) showed that IQ varied depending upon whether the same type of item was imbedded in easy or hard tests, and that subjects either adapted immediately or in successive stages as function of difficulty level. Dollin and Sakoda (1962) have just shown sequential effects of cards of a certain emotional tone in the Thematic Apperception Test upon the tone attributed to succeeding cards, as suggested by Helson's stimulus-background formulations.

Since order of stimuli appears to affect ALs within tests, the order in which tests are arranged in a battery may affect one another too; that is, each test interacts with the others as a series stimulus at the same time serving as a background experience for the next test confronted by the subject. Given this view, the clinician should be alerted to the necessity of charting the vicissitudes of the subject's shifts from one task to another as revealing overall adaptation.

#### IMPLICATIONS OF AL THEORY FOR THE CLINICAL TEST SITUATION

There are three implications of the theoretical model proposed herein for the clinical test situation:

*Theoretical.* The theory of AL offers a scientific explanation in place of a quasi-mystical description of the clinical test situation. It substitutes an operational, quantifiable framework for makeshift explanations. Chiefly, it provides the necessary relativity to account for the adaptive behavior of the subject rather than an absolutist view of fixed trait behavior. The theory of AL opens

the way to a further understanding of the process of clinical judgment heretofore largely refractory to scientific study.

It also offers the possibility of new insights into personality in adaptation terms, such as the meaning of integration and the psychological dynamisms impeding or facilitating new adaptations. From the standpoint of general psychology, AL theory provides rapprochement between psychology of the laboratory and psychology of the clinic. If it is a truism that increased mathematization is essential to the development of a science, then the way may be paved by AL theory for clinical psychology to take its place within this historical process.

*Experimental.* The theory of AL by virtue of its operational nature may be expected to stimulate and facilitate further experimentation on variables attending the clinical test situation. Thus the way is opened towards assessing the relative contributions of prior experience, background, and focal stimuli vis-à-vis subjects.

The terra incognita of interactions among tests in a battery is now open to exploration. Factors influencing clinical judgments can be investigated systematically with a view towards improved training. The problems of mensuration involved in clinical testing may be subject to scrutiny with a degree of precision approaching psychophysical experiments.

*Practical.* Advantages that may accrue to the clinician through the adaptation model are: (a) greater understanding of the adaptive capacities of the individual vis-à-vis different tasks and stresses; (b) ascertaining the degree of consistency of a subject—that is, what is generally typical of a subject as against what may be typical of a subject momentarily; (c) an objective index of adjustment or adaptation and a dynamic

formulation of the adaptive process;  
 (d) a common theoretical framework within which to view the patient's performance on a variety of different tasks;  
 (e) a further understanding of the process of clinical judgment towards improved training in use of instruments and in clinical role-taking behavior.

## REFERENCES

- BLOCK, W. E. A preliminary study of adaptation-level theory as a framework for projective testing. Paper read at American Psychological Association, St. Louis, 1962. (a)  
 BLOCK, W. E. Psychometric aspects of the Rorschach technique. *J. proj. Tech.*, 1962, 26, 162-172. (b)  
 DOLLIN, A., & SAKODA, J. M. The effect of order of presentation on perception of TAT pictures. *J. consult. Psychol.*, 1962, 26, 340-344.  
 HEIM, A. W. Adaptation to level of difficulty in intelligence testing. *Brit. J. Psychol.*, 1955, 46, 211-224.  
 HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.  
 HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 1. New York: McGraw-Hill, 1959. Pp. 565-621.  
 HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.*, 1960, 57, 116-131.  
 HUTT, M. L. A clinical study of consecutive and adaptive testing. *J. consult. Psychol.*, 1947, 11, 93-103.  
 MASLING, J. The influences of situational and interpersonal variables in projective testing. *Psychol. Bull.*, 1960, 57, 65-85.  
 MICHELS, W. C., & HELSON, H. A reformulation of the Fechner law in terms of adaptation-level applied to rating-scale data. *Amer. J. Psychol.*, 1949, 62, 355-368.  
 MURSTEIN, B. I. A conceptual model of projective techniques applied to stimulus variations with thematic techniques. *J. consult. Psychol.*, 1959, 25, 3-14.  
 RETHLINGSHAFFER, D., & HINCKLEY, E. D. Influence of characteristics of judges on their psychophysical judgments. *Amer. Psychologist*, 1954, 9, 454. (Abstract)

(Received November 3, 1962)

# ADAPTATION LEVEL, PERSONALITY THEORY, AND PSYCHOPATHOLOGY<sup>1</sup>

SANFORD GOLDSTONE

AND

JOYCE LEVIS GOLDFARB

*Houston State Psychiatric Institute, Texas*

*Baylor University College of Medicine*

Adaptation-level (AL) theory represents a quantitative approach to the concept of adjustment and offers an experimental psychological point of departure for a science of personal and social behavior. It is suggested that a personality theory develop from classification of response types specifying relevant aspects of the focal, background, and residual factors. The various psychopathologies may be similarly viewed in terms of specific behavioral excesses or impairments within the framework of AL theory. A proposal for the use of AL theory in regard to personality and psychopathology is outlined and representative studies are cited.

Somewhat more than a decade and a half has elapsed since Helson (1947) presented his adaptation-level (AL) formulation as a general approach to psychophysical data. The theory of AL might not have become more than an intramural psychophysical issue, if it did not expand into a model for frames of reference, a schema for a psychology of adjustment, and the embryo for a general behavior theory.<sup>2</sup> Exponents of this point of view studied the applicability of its generalizations about adjustment and pooling within psychophysical judgment areas (Michels & Helson, 1954), and then ventured, with success, into the fields of motivation and learning (Bevan & Adamson, 1960), personal and social variables (Blake, Mouton, & Hain, 1956; Blake, Rosenbaum, & Dur-

yea, 1955; Rosenbaum & Blake, 1955), clinical judgment (Campbell, Hunt, & Lewis, 1957), and intelligence (Heim, 1955).

This paper is specifically concerned with a second look at AL schema that emerged from a psychophysical tradition which is offering promise as a general approach to a science of the person and his psychopathologies at a time when personality theories are severe disappointments.

## PERSONALITY THEORY

Our psychophysical ancestors and their psychometric cousins did not foresee modest excursions into sensory and motor functions leading to a testing movement that would produce an arrogant step-grandson called clinical psychology demanding a recognized place in a respectable science while violating all of the ground rules. This step-grandson flourished and is presumably doing mankind a service but is only loosely tied to general psychology by a psychometric tradition, an increasing accumulation of unorganized and contradictory research results, fragile bridges between examining room observations and the psychologies of learning, sensation and perception, and fiat.

The clinical psychologist with his

<sup>1</sup> Paper presented in "Symposium: Contemporary Problems in Adaptation-Level Theory," American Psychological Association, St. Louis, 1962. Work supported by United States Public Health Service Grant M-1121.

<sup>2</sup> The interchangeable use of the words *schema* and *theory* reflects the authors' conflict concerning purity regarding these concepts. Without presenting any tedious discourse rationalizing either word within a philosophy of science, schema and theory will be used to satisfy all, while emphasizing the position that an adaptation-level formulation provides useful order and system at the same time that it measures.

functional concern with disorder rapidly outgrew the psychophysical tradition when the returns from an era of individual differences and mental testing diminished. Even the more sophisticated taxonomic factor analysts who attempted to provide clinical psychology with a more up-to-date psychometry failed to provide a shoe that fit the burgeoning clinical foot. Psychodiagnosis and psychotherapy became more complex than a reaction time or keenness of discrimination study, and even the higher mental functions of Binet were not high enough to deal with the problems of mankind that fell into the receptive laps of psychologists.

Clinical psychology with its mental tests emerged from a psychophysical heritage and then renounced its ancestral ties. With ineffective conceptual and methodologic models available in traditional laboratory psychology, the clinician was forced to take matters into his own hands, and out of his rank a new breed was born—the personality theorist. This specialist was created to develop a science of the person that would receive acceptance as theory and act as a clinically useful point of departure.

However, a science of the person should not reflect the desperation and impatience of the clinician, and the arts and sciences of behavior must meet within common frames of reference in accordance with the developmental level of each.

No aspect of a scientific psychology can defend itself as some off-brand humanism, or on-brand philosophy of existence.

Psychologists can learn much from humanists concerning missions and technologic goals. The humanist can remind the psychologist that one of his responsibilities is to regard human subjects as people.

The philosopher of existence and his

band of psychoanalytic followers can provide an occasional astute observation that might be incorporated into a scientific philosophy. In the maze of fanciful and literary excursions into the realm of mythology, or pronouncements about the destiny of man, whether optimistic or pessimistic, whether toward self-destruction or self-realization, one might find profound and testable observations, or more likely get lost and ensnared.

When one accepts the concept *personality theory*, one accepts a frontal assault upon *the person* that measures directly the integration of all of the attributes of cognition, conation, sensation and perception, motivation and learning, traits and dispositions that offer at once the common and unique qualities of a given intensity regarding an organism within a species. This would require a separate science of the person with its own methods and language.

Psychological attributes combine as the dimensions of behavior that interact or pool into the entity we casually call personality. No theory can ignore the obligation to identify and label these dimensions, measure their attributes, and explicate the process of pooling that leaves us with the person. The refusal to highlight a morphology of the person within a scientific behavior theory is inexcusable. The clinician's abhorrence of dissection and the sight of the psychological blood reflects his commendable humanism, but the systematic labeling of the gross and microscopic anatomy of the psychological cadaver may lead to an ordered science of the person and his behavior pathology.

What are now called theories of personality are not theories at all. They either represent philosophies of being, methods designed to study unsystematized behavioral attributes (Cattell, 1946; Eysenck, 1947; Sheldon, 1940, 1942, 1954), or some translation of a

philosophy of existence into a behavior theory with much lost in the translation (Dollard & Miller, 1950; Mowrer, 1950).

Psychology should acknowledge the utility of these discourses on the possible nature of people in the clinical arts. Orderly philosophies of man by various psychoanalysts or other-named psychotherapists are useful for the clinician who must describe observations in a communicable, literary way. These tempting philosophies are functional for the clinical psychologist as a language for descriptive psychodiagnostics and verbal psychotherapies. These fluent discourses are not theories and a science of man will not emerge from clinically useful literary expositions.

A revisit to the psychophysical neighborhood of his parents might be of more than sentimental value to the clinical psychologist.

#### PSYCHOPHYSICS, PERSONALITY THEORY, AND PSYCHOPATHOLOGY

The psychophysical methods which may be identified both with the emergence of scientific psychology and clinical psychology have at times been treated as ends in themselves, and the term *psychophysics* has been synonymous with classical procedures used to settle threshold and difference limen problems of interest to very few people.

A "simple" absolute threshold determination is variable, relative to a variety of contextual and organismic factors. Judgment is a consequence of a pooling process which integrates the patterns of focal and background inputs, as well as concepts or past experience residuals based upon meetings with similar stimuli. Present psychophysical laws and their mathematical accompaniments provide rough approximation predictions within a restricted domain of laboratory circumstances. The tradition that

started with Weber and Fechner had a more noble goal that sought to relate the responses of organisms to the energetic configurations imposed upon their receptors. Any general approach to psychophysics represents a study of the judgment process under the plethora of available focal and background input conditions, and the endless varieties of organismic states.

This less than modest goal of creating a general theory, or a scientifically useful conceptual schema for the clinician and psychopathologist out of a psychophysics that appeared a century ago to study the sensitivity of the organism, may appear to represent a regressive reversion to the unsuccessful attempts by Kraepelin (1896) to use the psychophysical laboratory as the prototype for a scientific psychiatry. The difference rests with the functional distinctions between classical and frame of reference psychophysics. The latter permits a behavior taxonomy based upon responses to focal and background inputs, as well as classification of the attributes of concepts.

#### AL AS A GENERAL CONCEPTUAL SCHEMA

The clinical psychologist created the notions of personality and personality theory in the quest of a functional psychology of the integrated dimensions of behavior that produced an ordered or disordered person. Although their concern was with adjustment and adaptation, their holistic, humanistic philosophies distracted them with an obsessive concern with motivation and purposes of mankind.

Every psychophysical experiment is a study of adjustment. Any response to a configuration of focal and background inputs with a given organismic state represents an adaptation. Although obvious, devotion to method blinded the psychophysicist to his calling as a gen-

eral behavior theorist of adaptation. It remained for Helson (1947, 1948, 1955, 1959) to recognize the relationship between specific and general instances of adaptation; a conceptual schema as a point of departure for a quantitative theory of behavior was established. Others recognized the importance of focal and background conditions in scale formation and the designation of indifference points. The homeostatic emphasis in physiologic (Cannon, 1939) and behavioral contexts is not new. However, with the rallying concept *adjustment* Helson brought these considerations together into a single systematic schema that permits manipulation of contextual and residual variables, measurement, and first approximation mathematical formulations.

Although AL theory began as a series of generalizations regarding spectral energy input and background reflectances on the one hand, and perceived color on the other (Helson, 1938; Helson & Jeffers, 1940), its author (Helson, 1947, 1948) took the risk of criticism for extending a laboratory and physiologically based concept *adaptation* through analogy. Is the sensory psychological and sensory physiological time-honored identification of adaptation related to adaptation within a series of lifted weights, or are we dealing with a semantic artifact which is a function of the labels within a category scaling procedure (Stevens, 1958)? Does personal adaptation so basic for the clinician bear any resemblance to the spirit and letter of these terms as understood in the laboratory? Is this extension of adaptation from its physiological to absolute judgment to personal and social contexts little more than a word game?

It is well known that AL schema (Helson, 1947, 1948) defines adjustment in terms of responses to situations scaled along bipolar continua indicating neu-

tral, or balanced states. This formulation viewed the neutral AL as a weighted geometric mean of three classes of stimuli affecting an organism within a temporal context: the inputs primarily within the contemporary, focal scene; all other inputs imposed upon available receptors forming the background for the focal circumstance; all organismic factors including residuals or concepts based upon historic encounters with relevant inputs, as well as constitutional and dispositional factors. All three are hopefully definable in stimulus terms and pool through an integration process to produce AL.

The specific details and assumptions of AL theory regarding the judgment process as well as its dangerous but successful missions beyond the psychophysical bench are reviewed elsewhere (Allport, 1955; Guilford, 1954; Helson, 1947, 1948, 1955, 1959). While Helson's psychophysical upbringing leaves him with less interest in the third category of pooled stimulation, residuals or concepts, he acknowledges (Helson, 1959) that herein rest the critical areas for personality theory. The student of healthy and pathologic personal behavior is concerned with the development of residuals, the mechanics of their pooling with contemporary stimulation, and the process of conceptual modification by the manipulation of focal and background conditions to facilitate effective adjustment to changing internal and external contexts (Helson, 1955, 1959).

#### AL AND PERSONALITY

Sympathy with global theories of personality allows the bypassing of fundamental psychological attributes and dimensions that integrate or pool leaving a "whole, undissectable person." This may appear as a short cut but it is unlikely, after a half-century of floundering, that a general behavior theory of the

person will emerge from the examining room or mental health agencies. It is probable that the lessons of history will point to the laboratory for a science of people.

If this is the case how can the clinician avoid losing the person? How can AL schema provide methods and systems that will not violate the person, and his healthy or psychopathologic adaptation?

There are two global methods to directly approach personality and personality theory: those involving the mechanics of proof emphasizing statistical design and rigor, or others operating through clinically useful but scientifically deplorable literary expositions (Adler, 1954). Neither approach provides scientifically useful theories, models, or schemata.

There are two approaches to the application of AL theory to the science of the person. One involves a direct attack upon attitudes and everyday behavior dynamics using the AL model in a study of what is generally called personality and social variables. Investigations of volunteering and generosity (Blake, Rosenbaum, & Duryea, 1955; Rosenbaum & Blake, 1955) using laboratory derived principles of judgment represent a direct application of AL schema to personal and social situations.

A second keeps the investigator and theorist in the laboratory and although he, too, manipulates focal and background conditions as well as the more elusive residuals, he stays with well-graded input continua emphasizing traditional magnitude and sensory dimensions.

This is simply a difference in research strategy. The direct application of AL principles to personal and social variables has been the more usual approach when this schema has been extended to include the behavior of people in every-

day life situations. However, one should not overlook the possibility that a theory of personality and psychopathology might just as well emerge from a classification of response types based upon traditional laboratory procedures.

#### BIOMETRICS AND AL

Adjustment within a given series of weights associated with a controlled background context along with known organismic characteristics is as much an experiment in, and measurement of personality, as some crystal ball that reveals hidden hostilities to poor unsuspecting siblings, or ghosts from the past awaiting exorcism upon the couch. Category judgments of duration within a well-graded temporal series under specific sensory and anchoring conditions can be considered as much a measure of personality as the judgment of the subject who may or may not volunteer in the face of varied social pressures. All of these are measures of personal behavior, but a taxonomy that may permit a personality theory is more likely to emerge from the laboratory.

Surely all psychophysical responses are partly if not primarily cognitive. Investigators need only seek modal responses to specific focal and background conditions, within the framework of specific organismic or residual states, and define these modal responses as the dimensions of behavior to be systematized into a science of the person. With our present vague definitions of personality, no one will deny that a laboratory psychophysical adjustment is a measure of personality, but the clinician will ask about the relationship between adaptation to a series of sounds of varied intensity and adjustment to everyday life experiences.

Simple laboratory adjustments to series of magnitudes are indeed adjustments to everyday life situations. The

question is not one of general relevance, but of specific relevance. The scaling of intensity, extensity, or protensity is part of general adjustment. We are all exposed to everyday life analogues of the various psychophysical methods and scaling procedures within the framework of a multitude of dimensions. These judgmental circumstances may, with their contextual variations, encompass the gamut of problems faced by clinicians in the examining room.

Zubin's (Burdock, Sutton, & Zubin, 1958; Zubin, 1958) close association with problems of personality and psychopathology has increased the firmness of his position that regards the psychology of ordered and disordered people as a laboratory science. Proposing a biometric approach to the study of personality and psychopathology, he offers five relevant, measurable activity levels that may be investigated quantitatively during a base line idling state and under stress-loading conditions. Zubin (1958) courageously turned back the clock and re-emphasized the objective techniques offered through psychophysics and psychometrics. His biometric approach hopes to integrate the data from the hierarchy of the five basic functioning levels: physiologic, sensory, perceptual, psychomotor, and conceptual under base line and stressful conditions. This view requires a laboratory approach to personality and psychopathology focusing upon the integration of conditions of stimulation and levels of organismic functioning, pleading for rallying concepts and models that exploit consistencies and discover major sources of variance.

A biometric model requires a taxonomy of behavior based upon the psychophysical and psychophysiological sciences and their relationships. Zubin's search for schematization reflects a concern about the undirected multiplication

and accumulation of facts and pathology; his (Burdock, Sutton, & Zubin, 1958) dismay with current conceptions of personality and personality theory leads to the conclusion that one cannot take for granted any functional interdependence regarding these theories and mental disease. This is not the case if one begins with a biometric AL approach to personality and psychopathology.

Biometrics provides a useful empirical track and offers wise suggestions regarding the need for schematization. However, the approach of biometricians is first and foremost a devotion to measurement and the classification of broad areas to be measured. The homeostatic issue of adjustment within AL theory may provide the schema for ordering the multitude of biometric facts that are accumulating.

#### AN AL TAXONOMY OF PERSONALITY AND PSYCHOPATHOLOGY

A psychological classification system based upon a quantifiable concept of adjustment derived from AL theory may permit a scientific psychology of the person and his psychopathology. The characteristics of input signals that convey focal and background information determining adjustment levels for people or groups can be specified with precision and accuracy. It is feasible to scale individuals according to response types based upon characteristic modes of behaving in regard to controlled foreground and background conditions. Individuals differ with reference to the extent to which they weight or pool various aspects of an external context; subject share certain common response patterns within a specific external context; individual differences exist regarding relevance of details within the focal background contexts; people do carry equivalent residuals from

experience into experimental and clinical contexts; and, just as personality may be defined in terms of AL response types, so can specific excesses be classified as psychopathology.

Personality and psychopathology can be defined in terms of ALs in relation to specific focal, and background contexts on the one hand, and organismic residuals on the other.

It will be necessary to expose the seemingly endless relevant contextual factors in a psychophysical judgment, quantitatively assess their contribution to the AL, and apply derived principles across dimensions to evaluate their generality. We can point to central tendency effects, time-order-errors, sequential effects, end effects, contrast and assimilation effects, and ALs located in the neighborhood of the weighted geometric mean of a series. These offer first approximation generalizations that are dependent upon an abundance of known and unknown contextual and residual factors. With AL so dependent upon the capricious and very complex context, is it any wonder that contradictory findings appear regarding specific aspects of judgment?

The broad schematization has proven useful; a pooling theory of adjustment should be fruitful for a general psychology of the person. Now it is necessary to get down to the business of the exploration of details. The microscopic components of a broad schema represent its experimental essence. A broad schema that leads to a point where one must abandon hope for quick theory, but does lead to a point of increasing experimental returns requiring effortful exploration of detail is likely to leave one on a productive road. The AL is ready for the step beyond broad, schematic generalizations.

1. There is the tedious task of identifying the many relevant attributes

within a given judgmental dimension that contribute to an adaptation level. This will be a long-suffering, unglamorous enterprise. It is known that sequential effects (Parducci, 1962) and channel of input (Behar & Bevan, 1960, 1961; Goldstone, Boardman, & Lhamon, 1959; Goldstone & Goldfarb, in press; Lhamon, Edelberg, & Goldstone, 1962) are relevant in specific judgmental circumstances. Absolute threshold is not a limiting factor in scale shifting regarding intensity and size judgments (Black & Bevan, 1960; Boardman & Goldstone, 1962; Goldstone, Goldfarb, Strong, & Russell, 1962), stimulus range, and presentation interval influence judged loudness (Bevan, Reed, & Pritchard, 1961), and information regarding relevant aspects of effective anchors in form judgment is now available (Bevan & Pritchard, 1963). A long list of stimulus and presentation parameters is waiting to be investigated in seemingly endless combinations to determine what influences judgment.

2. After relevance within a judgment dimension is established and classified, the measurement of the extent of influence is a next step. Concern with the amount of impact of a stimulus parameter under a controlled presentation circumstance, the effect of a presentation circumstance under controlled conditions of stimulation and the interaction of stimulus and presentation variables are of basic interest.

3. It is important to know whether a relevance which produces predictable alterations in the AL is specific to a given judgmental dimension, or if one possesses a more general principle of behavior. It is necessary to derive measurable similarities and differences in the effect of a contextual factor across judgment dimensions.

4. The elusive residual may be viewed as the wastebasket for AL theory or

it may be an inescapable and even desirable region for study. Traditionally the psychophysicist focuses upon the contemporary scene and the residuals or concepts learned from previous encounters with relevant stimuli are ignored, trained-out, or averaged into oblivion through the use of groups. Insofar as these concepts reflect response dispositions they are of essential importance to the student of the person. Insofar as residuals represent the immediate displays of disorder they are the primary concern of the psychopathologist. A science of the person and his psychopathology is concerned less with the organism as a "null instrument" and more with perceptual-conceptual (Brunner, 1957; Goldstone, 1962) behavior grossly classified in terms of attitudes, thinking, and experience. The absolute method emphasized by AL theory represents a more relevant laboratory analogue of the personal encounters involved in everyday life adjustments. Whereas a comparison method provides an external standard to an organic meter, the absolute method requires the person to pool input configurations with internal norms or concepts. Controlled focal and background configurations reveal data concerning the status of these internal standards. The behavioral dysfunctions of the various psychopathologies (Goldstone, 1962; Zubin, 1958) are rarely of a metering or sensitivity nature. Disordered absolute judgments indicate alterations, impairments, or deficiencies in the scaling process or the residual concept. The student of the person and his mental disorders is particularly interested in the use of past experience and the ability of concepts to pool with relevant contemporary inputs and ward off inappropriate background noise.

Residuals are learned and are available within certain acceptance-rejection limits. Inputs are compared with these

concepts and a neutral point is established.

The delusion and hallucination of the psychotic, the confusion of the toxic state, the exaggerations or denials of the neurotic, and the ineffectiveness accompanying personality disorder and mental defect all reflect absolute judgments that do violence to either the nature of an accepted norm or the accepted range of that norm . . . [Goldstone, 1962, p. 264].

The theory of AL permits the definition of a person's disordered judgment in terms of his specific response to the detailed focal or background stimulus configurations. Conceptual impairment may be a function of insufficient learning, incorrect learning, conflictual learning, or, a contemporary disruption in conceptualization and scaling due to such other organismic factors as fatigue, intoxication, or mental illness.

These residuals need not be considered mysterious aspects of the conscious or unconscious mind unapproachable through laboratory techniques. Concepts may be assigned attributes much like the sensations. These attributes such as vividness or intensity, availability, range of identity and range of response categorization (Goldstone, 1962) may be investigated by developmental studies of learned social norms, as well as by studies of immediate and successive pooling in order to explore the persistence and strength of these residuals in the face of relevant and irrelevant focal and background stimulation.

The taxonomy of these residuals and their pathology remain as a challenge to the AL theorist.

5. As each relevant internal and external influence is discovered, labeled, and extensively explored within and beyond a specific judgment dimension, the theorists may begin moving from methodologic successes to theory. Increasingly sophisticated measurement and the discovery of additional relevant factors and their mode of operation may direct

the pooling schema of AL to a general behavior theory of adjustment. The definition and measurement of more attributes of focal, background, and residual conditions will complicate the mathematics of a psychophysical law but the principle of adjustment will remain unchanged. The process of pooling, patterning, or organization will become the scientific focus of attention. It is now essential to determine what is being pooled and the conditions under which a given quality and quantity of pooling transpires.

The need for a functional biometric, AL model for personality and psychopathology cannot be overemphasized. The clinical-experimental psychologist is faced with the problem of ordering the abundance of research reports that accumulate daily offering generalizations about people and mental disease. It becomes more perplexing when people with schizophrenia are compared with healthy controls and found to have reduced size constancy, overconstancy, or equivalent constancy (Rausch, 1952; Reynolds, 1953; Sanders & Pacht, 1952; Weckowicz, 1957). The reduced constancy accompanying schizophrenia obtained by Boardman, Goldstone, Reiner, and Himmel (1964) was interpreted in terms of specific conceptual alterations accompanying the illness, in relation to specific contextual factors.

More confusion is available when one notes that with identical psychophysical methods, schizophrenic patients show increased overestimation of 1-second durations (Lhamon & Goldstone, 1956; Weinstein, Goldstone, & Boardman, 1958; Wright, Goldstone, & Boardman, 1962) while underestimating the length of 1 inch (Boardman, Goldstone, Reiner, & Fathauer, 1962). Healthy subjects were more accurate and precise, and, less influenced by anchors, when making absolute judgments of an inch than of

a second (Boardman, Aldrich, Reiner, & Goldstone, 1959; Goldstone, Lhamon, & Boardman, 1957). Salzinger (1957) found increased anchor effects by schizophrenic subjects in a weight-lifting experiment which was confirmed in regard to spatial judgments (Boardman, Goldstone, Reiner, & Fathauer, 1962). This was not found with similarly obtained temporal judgments (Webster, 1962). General statements about magnitude judgment in general and relations with personal or psychopathology variables must consider the specific dimension or sensory attribute and likely relevant contextual influences.

Even general statements about magnitude judgments by healthy and psychopathologic subjects are not safe within a given dimension. It was discovered that the generalization that schizophrenic patients overestimated auditory durations obtained from three replications (Lhamon & Goldstone, 1956; Weinstein, Goldstone, & Boardman, 1958; Wright, Goldstone, & Boardman, 1962) was an oversimplification. The alteration was not a general property of time judgment at all, but a function of specific aspects of the temporal circumstance. The original studies used a modified method of limits procedure with an arithmetic series. When the stimulus arrangements were altered to either a geometric step interval (Wright, Goldstone, & Boardman, 1962) or the constant method (Webster, Goldstone, & Webb, 1962) the schizophrenia-healthy differences disappeared. Close attention to these important contextual details and their differential effect upon various populations will provide the basic information about healthy and psychopathologic response types. Contextual influences upon patient-control differences are not sources of experimental error, but represent the essence

of these differences and provide a basis for a personal taxonomy.

Other studies have shown reduced successive pooling by schizophrenic patients in time-judgment experiments using alternate anchors (Webster, 1962); Weinstein, Goldstone, & Boardman, 1958) and channel of input (J. L. Goldfarb, unpublished), and, sense mode has also been suggested as a relevant dimension in predicting patient-control differences (Wright, Goldstone, & Boardman, 1959). Impaired successive pooling and reduced intersensory discrimination accompanying schizophrenia permitted a psychological explanation of hallucinations based upon the AL schema (Goldstone, 1962).

Pointless arguments about contradictory results (Aronson, Silverstein, & Klee, 1959; Boardman, Goldstone, & Lhamon, 1957) do little more than emphasize the oversimplification involved in general statements about personal or group characteristics when the relevant momentary input and internal details are not considered. There are no contradictory or conflicting results, only insufficient attention to, and specification of, the relevant focal, background, and residual factors.

Regarding the context alone, AL followers have a long, hard road cut out for them. Adding a scientific concern for the less manageable residuals complicates but completes a far-reaching goal of a science of the person and his psychopathology.

#### SUMMARY

The direct, global assaults upon a possible science of the person have provided us with useful clinical philosophies but have not left us with conceptual approaches that might resemble embryonic theory.

As a quantitative approach to adjustment, the broad AL schema offers a pos-

sible general psychological point of departure that might lead to a theory of personality.

It is proposed that a laboratory based AL theory of personality develop from a classification of response types specifying the relevant aspects of the focal, background, and residual factors.

The various psychopathologies may be similarly viewed in terms of the specific excesses or impairments within the framework of Helson's AL and Zubin's biometric models.

Every factor relevant to the location of a neutral point and every influence that shifts a reference scale is an attribute of personality. It is part of the person and he may be compared with others, or at times, with himself, with regard to these factors and influences. If there is to be a science of the person or a personality theory, every adjustment determined by the abundance of momentary properties of focal and background stimulation, as well as the momentary residuals or state of the organism must be considered a personality measure. Whether or not a specific laboratory measure of adjustment is relevant to the personal adjustments of interest to the clinical psychologist is for the scientific psychopathologist to decide. This newer breed of personality theorists, emerging from both the laboratory and the examining room can designate the dimensions of adjustment that are of interest to the clinician. These can be specified in the language of AL and the focal and background inputs can be manipulated in the laboratory under idling and stressful conditions developing a taxonomy of personality which will permit comparison of psychopathology with health.

#### REFERENCES

- ADLER, D. L. Some recent books on personality. *Psychol. Bull.*, 1954, **51**, 283-296.  
ALLPORT, F. H. *Theories of perception and the*

- concept of structure. New York: Wiley, 1955.
- ARONSON, H., SILVERSTEIN, A. B., & KLEE, G. D. Influence of lysergic acid diethylamide (LSD-25) on subjective time. *Arch. gen. Psychiat.*, 1959, 1, 469-472.
- BEHAR, I., & BEVAN, W. Analysis of the prime psychophysical judgment. *Percept. mot. Skills*, 1960, 10, 82.
- BEHAR, I., & BEVAN, W. The perceived duration of auditory and visual intervals: Cross modal comparison and interaction. *Amer. J. Psychol.*, 1961, 74, 17-26.
- BEVAN, W., & ADAMSON, R. Reinforcers and reinforcement: Their relation to maze performance. *J. exp. Psychol.*, 1960, 59, 226-232.
- BEVAN, W., & PRITCHARD, JOAN F. The anchor effect and the problem of relevance in the judgment of shape. *J. gen. Psychol.*, 1963, 69, 147-161.
- BEVAN, W., REED, W. G., & PRITCHARD, JOAN F. Single-stimulus judgments of loudness as a function of presentation-interval. Technical Report No. 3, 1961, Nonr-3290 (01), Project NR 142-155, Office of Naval Research.
- BLACK, R. W., & BEVAN, W. The effect of subliminal shock upon the judged intensity of weak shock. *Amer. J. Psychol.*, 1960, 73, 262-267.
- BLAKE, R. R., MOUTON, JANE S., & HAIN, J. D. Social forces in petition signing. *SW soc. sci. Quart.*, 1956, 36, 385-390.
- BLAKE, R. R., ROSENBAUM, M., & DURYEA, R. A. Gift giving as a function of group standards. *Hum. Relat.*, 1955, 8, 61-73.
- BOARDMAN, W. K., ALDRICH, R. C., REINER, M. L., & GOLDSTONE, S. The effects of anchors on apparent length. *J. gen. Psychol.*, 1959, 61, 45-49.
- BOARDMAN, W. K., & GOLDSTONE, S. Effects of subliminal anchors upon judgments of size. *Percept. mot. Skills*, 1962, 14, 475-482.
- BOARDMAN, W. K., GOLDSTONE, S., & LHAMON, W. T. Effects of lysergic acid diethylamide (LSD) on the time sense of normals. *Arch. Neurol. Psychiat.*, 1957, 78, 321-324.
- BOARDMAN, W. K., GOLDSTONE, S., REINER, M. L., & FATHAUER, W. R. Anchor effects, spatial judgments, and schizophrenia. *J. abnorm. soc. Psychol.*, 1962, 65, 273-278.
- BOARDMAN, W. K., GOLDSTONE, S., REINER, M. L., & HMMEL, S. Constancy of absolute judgments of size by normals and schizophrenics. *J. abnorm. soc. Psychol.*, 1964, 68, in press.
- BRUNER, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, 64, 123-152.
- BURDOCK, E. I., SUTTON, S., & ZUBIN, J. Personality and psychopathology. *J. abnorm. soc. Psychol.*, 1958, 56, 18-30.
- CAMPBELL, D. T., HUNT, W. A., & LEWIS, N. A. The effects of assimilation and contrast in judgment of clinical materials. *Amer. J. Psychol.*, 1957, 70, 347-360.
- CANNON, W. B. *The wisdom of the body*. New York: Norton, 1939.
- CATTELL, R. B. *Description and measurement of personality*. New York: World Book, 1946.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- EYSENCK, H. J. *Dimensions of personality*. London: Routledge & Kegan Paul, 1947.
- GOLDSTONE, S. Psychophysics, reality and hallucinations. In L. J. West (Ed.), *Hallucinations*. New York: Grune & Stratton, 1962.
- GOLDSTONE, S., BOARDMAN, W. K., & LHAMON, W. T. Intersensory comparisons of temporal judgments. *J. exp. Psychol.*, 1959, 57, 243-248.
- GOLDSTONE, S., & GOLDFARB, JOYCE L. Auditory and visual time judgment. *J. gen. Psychol.*, in press.
- GOLDSTONE, S., GOLDFARB, JOYCE L., STRONG, J., & RUSSELL, J. Replication: Effect of subliminal shock upon judged intensity of weak shock. *Percept. mot. Skills*, 1962, 14, 222.
- GOLDSTONE, S., LHAMON, W. T., & BOARDMAN, W. K. The time sense: Anchor effects and apparent duration. *J. Psychol.*, 1957, 44, 145-153.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- HEIM, ALICE W. Adaptation to level of difficulty in intelligence testing. *Brit. J. Psychol.*, 1955, 46, 211-224.
- HELSON, H. Fundamental problems in color vision: I. The principle governing changes in hue, saturation, and lightness of non-selective samples of chromatic illumination. *J. exp. Psychol.*, 1938, 23, 439-476.
- HELSON, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, 60, 1-29.
- HELSON, H. Adaptation-level as a basis for quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.
- HELSON, H. An experimental approach to personality. *Psychiat. res. Rep.*, 1955, 2, 89-99.
- HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. I. New York: McGraw-Hill, 1959.
- HELSON, H., & JEFFERS, V. B. Fundamental problems in color vision: II. Hue, lightness, and saturation of selective samples in chrom-

- matic illumination. *J. exp. Psychol.*, 1940, 26, 1-27.
- KRAEPELIN, E. Der psychologische Versuch in der Psychiatrie. *Psychol. Arbeit.*, 1896, 1, 1-91.
- LHAMON, W. T., EDELBERG, R., & GOLDSTONE, S. A comparison of tactile and auditory time judgment. *Percept. mot. Skills*, 1962, 14, 366.
- LHAMON, W. T., & GOLDSTONE, S. The time sense: Estimation of one-second durations by schizophrenic patients. *Arch. Neurol. Psychiat.*, 1956, 76, 625-629.
- MICHELIS, W. C., & HELSON, H. A quantitative theory of time-order effects. *Amer. J. Psychol.*, 1954, 67, 327-334.
- MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
- PARDUCCI, A. Sequential effects in judgment. Paper read at American Psychological Association, St. Louis, 1962.
- RAUSCH, H. L. Perceptual constancy in schizophrenia: Size constancy. *J. Pers.*, 1952, 21, 176-187.
- REYNOLDS, G. A. Perceptual constancy in "schizophrenics" and "normals." Unpublished doctoral dissertation, Purdue University, 1953.
- ROSENBAUM, M., & BLAKE, R. R. Volunteering as a function of field structure. *J. abnorm. soc. Psychol.*, 1955, 50, 193-196.
- SALZINGER, K. Shift in judgment of weights as a function of anchoring stimuli and instructions in early schizophrenics and normals. *J. abnorm. soc. Psychol.*, 1957, 55, 43-49.
- SANDERS, R., & PACT, A. R. Perceptual size constancy in known clinical groups. *J. consult. Psychol.*, 1952, 16, 440-444.
- SHELDON, W. H. *The varieties of human physique: An introduction to constitutional psychology*. New York: Harper, 1940.
- SHELDON, W. H. *The varieties of temperament: A psychology of constitutional differences*. New York: Harper, 1942.
- SHELDON, W. H. *Atlas of men: A guide for somatotyping the adult male at all ages*. New York: Harper, 1954.
- STEVENS, S. S. Adaptation-level vs. the relativity of judgment. *Amer. J. Psychol.*, 1958, 71, 633-646.
- WEBSTER, F. R. Studies in time perception: Contextual factors and psychopathology. Unpublished doctoral dissertation, Vanderbilt University, 1962.
- WEBSTER, F. R., GOLDSTONE, S., & WEBB, W. W. Time judgment and schizophrenia: Psychophysical method as a relevant contextual factor. *J. Psychol.*, 1962, 54, 159-164.
- WECKOWICZ, T. E. Size constancy in schizophrenic patients. *J. ment. Sci.*, 1957, 103, 475-486.
- WEINSTEIN, A. D., GOLDSTONE, S., & BOARDMAN, W. K. The effect of recent and remote frames of reference on temporal judgments of schizophrenic patients. *J. abnorm. soc. Psychol.*, 1958, 57, 241-244.
- WRIGHT, D., GOLDSTONE, S., & BOARDMAN, W. K. Intersensory comparisons of temporal judgments by schizophrenic and healthy subjects. *Amer. Psychologist*, 1959, 14, 404. (Abstract)
- WRIGHT, D., GOLDSTONE, S., & BOARDMAN, W. K. Time judgment and schizophrenia: Step interval as a relevant contextual factor. *J. Psychol.*, 1962, 54, 33-38.
- ZUBIN, J. A biometric model for psychopathology. In *Current trends in the description and analysis of behavior*. Pittsburgh: University of Pittsburgh Press, 1958.

(Received November 18, 1962)

## RETENTION CURVES:

### FACTS OR ARTIFACTS? <sup>1</sup>

HARRY P. BAHRICK

*Ohio Wesleyan University*

Retention curves based upon recognition scores may be comparable in slope and amount of retention to curves based upon recall and anticipation performance. Previous contrary conclusions are overgeneralized. They are the result of the selection of easy recognition tests and failure to control the variable of overlearning. Measures of recognition, free recall and anticipation are dichotomous, and the slopes of curves based upon such measures are artifacts of the changing sensitivity of the measure. The curves can therefore not provide the basis for general conclusions regarding relative amounts of forgetting over different time periods.

Ambiguity regarding the interrelations among commonly used response indicators creates serious methodological as well as theoretical problems. If these interrelations are not well understood, overgeneralizations and misinterpretations of experimental findings are unavoidable (Bahrick, Fitts, & Briggs, 1957). The classic study establishing the interrelations among measures of retention was performed by Luh (1922). Postman and Rau (1957) recognized certain inadequacies of Luh's method. They collected extensive new data and thoroughly reviewed the literature on this topic. One of the principal differences between the findings of Luh and of Postman and Rau pertains to the curves based upon recognition measures. Postman and Rau found practically no decline of recognition scores over a 2-day period when subjects originally learned the material to a criterion of one perfect anticipation trial. They report: "For all practical purposes the method of recognition shows no forgetting at all during the time interval used in the present study [Postman & Rau, 1957, p. 291]." They suggest that the decline of recognition

performance over the same period reported by Luh was the result of methodological errors of interpolating tests of reproduction, as well as other uncontrolled interference effects which tended to lower subsequent performance on recognition tests. In regard to the interrelation between indicators of recall and recognition Postman and Rau (1957) concluded: "The one fact for which there is substantial experimental evidence is that tests of recognition yield higher scores than do tests of recall [p. 218]." This conclusion has extensive documentation (e.g., Achilles, 1920; Clarke, 1934; Hollingworth, 1913; Miler, 1957; Myers, 1914; Stalnaker, 1935). No comparable support exists for Postman and Rau's finding that recognition shows no retention loss at all over a 2-day period, but it is generally held that recognition curves show a more gentle slope and do not exhibit the negative acceleration observed in the classical recall and anticipation retention curves (Luh, 1922; Miler, 1957).

It is the principal purpose of this article to show that both of the above conclusions regarding the superiority of recognition to recall performance, and regarding the slope of retention curves are overgeneralizations, and therefore misleading, because the findings on which they are based do not represent

<sup>1</sup> This research was supported by United States Public Health Service Grant MH-05685-02.

The writer wishes to thank Louis Herman for his many helpful comments.

intrinsic differences between indicants of recognition and recall. Rather, the findings reflect the exclusive use of experimental designs which fail to control the variable of overlearning, the exclusive use of "easy" recognition tests in which correct and incorrect alternatives are highly differentiated, and artifacts arising from the use of dichotomous scoring procedures.

#### EXPERIMENTAL DESIGNS AND OVERLEARNING

In most studies of retention, stimulus material, rather than retention measures per se, have been the subject of investigation. Two general types of experimental design are conventionally used when retention of different stimulus material is investigated: (a) an equal amount of training is given to subjects in the various conditions to be compared and (b) subjects are trained to a comparable criterion under the various conditions. To illustrate the advantage of using *both* types of design let us suppose it is desired to compare retention for nonsense syllables high and low in association value. Following Design *a* an equal number of training trials is administered to the groups learning the two kinds of material, and retention is tested after suitable intervals. Following Design *b* the two groups are trained until they reach a comparable response criterion, such as the first trial on which all responses are correct. If only Design *a* is used, and it is found that retention is better for the syllables high in association value, it is not possible to determine if this is true just because the material is "easier," and was therefore learned to a higher degree. By supplementing this comparison with data from Design *b* one can determine whether the high association value affected retention, independent of the degree of original learning.

Investigations in which comparisons of retention measures rather than of stimulus material are the subject of interest have made use of two variations of Design *a*, but have failed to supplement this with information from Design *b*. In the traditional version of Design *a* investigators have given all subjects a constant number of training trials and have later compared performance on recognition and recall tests. In the second version all subjects are trained to the same criterion, usually the first trial of correct anticipation, followed by the comparison between recall and recognition performance on later retention tests. At first blush it would appear that this second version corresponds exactly to the design of Type *b*, and should therefore yield the same kind of information, but this is not so.

To compare retention measures by Design *b*, subjects of various groups must be trained until they have reached a comparable criterion on the *same* task on which retention is tested later. This, of course, entails administering different numbers of training trials for the various groups, similar to the case described earlier in which different numbers of training trials were administered to the groups learning nonsense syllables high or low in association value. In the second version of Design *a* described above, the subjects in the recall and recognition groups receive *equal* amounts of training, and are not brought to a comparable criterion on the recognition and recall tasks on which retention is later measured. Thus, the degree of original learning with respect to the condition being tested during retention has not been controlled. As a consequence this type of study does not yield the information usually available from designs of Type *b*, and as a matter of fact, does not add much to the information already

available from the traditional version of Design *a*.

To compare retention for recall and recognition by Design *b*, it is necessary to train one group of subjects until all of their recall responses are correct, and another group of subjects until all of their recognition responses are correct. The general conclusion that tests of recognition yield higher scores than do tests of recall has not been based upon any such comparison and is, therefore, limited to situations in which the degree of original training is not comparable. Because investigators have used only those recognition tests in which correct and incorrect alternatives were highly differentiated, the degree of learning with respect to the recognition task was generally much higher than the degree of learning with respect to the recall task when training was interrupted. The better performance on later retention tests, as well as the difference in slope of the retention curves based upon recall and recognition measures, reflects this difference in the degree of original learning. Thus, Postman, Jenkins, and Postman (1948) reported that after six trials of exposure to 48 nonsense syllables subjects were immediately able to make an average of 24.33 correct recognition responses, but only 7.20 correct recall responses if the recall test was administered first. They reported 27.66 and 9.09 correct responses, respectively, if the recognition test was taken first. Similarly, Achilles (1920) reported that subjects who were given 50 seconds to study a list of 25 nonsense syllables were immediately able to recall an average of only 12% of the list, but correctly recognized 42%. Hollingworth (1913) found that the number of trials needed to reach the threshold of recall was larger than the number of trials needed to reach the recognition threshold, and the ratio between the two was larger for meaning-

ful than for meaningless material. Such data have led Andreas (1960, p. 447), Miler (1957), and others to conceptualize a hypothetical threshold of recognition which is lower than a hypothetical threshold of recall, so that the strength of a given memory trace may be below the recall threshold, but above the recognition threshold.

It has long been known (Ebbinghaus, 1913) that material which is overlearned will be retained better, or, more precisely, that indicants of retention are not sensitive to early retention losses if the material has been overlearned with respect to the threshold of that indicant. Thus, when material has been sufficiently overlearned with respect to the recall criterion, recall scores may continue for long periods of time near the 100% correct level. We cannot demonstrate a classical retention curve by the recall measure for the words in the Lord's Prayer or the National Anthem for most adult Americans. A somewhat analogous situation may occur when performance on easy recognition tests shows little or no decline for several days after associations have been learned to the threshold of anticipation. This does not serve as a valid basis for contending, however, that recognition measures, per se, do not yield classical, that is, negatively accelerated, retention curves. The effect of the variable of overlearning upon the slope of recognition curves is apparent if we examine results from those investigations in which the degree of overlearning was more limited.

Luh also trained subjects to lesser criteria, such as 67% and 33% of the trials needed for the threshold of anticipation. He obtained recognition curves which showed significantly more loss during the early postlearning interval.

Strong (1913) reported a negatively accelerated curve for recognition scores

in a situation in which training seems to have been interrupted near the threshold of recognition. His data also tend to support the present view that failure to obtain substantial decline of recognition scores in the immediate postlearning interval is a result of overlearning. A more precise interpretation of his curves is difficult, however, because training was not interrupted precisely at the threshold of recognition, and because additional variables were introduced by requiring his subjects to indicate various degrees of confidence concerning their recognition responses. The negatively accelerated curve was obtained for the responses made with the greatest degree of confidence.

The generally accepted conclusion that performance on recognition tests is less affected by retroactive inhibition than performance on recall tests (Britt, 1935; Hollingworth, 1913; Postman & Rau, 1957) is also limited by the above considerations. Overlearning makes material less vulnerable to interference effects, and therefore needs to be controlled for a valid comparison of such effects. The studies on which such comparisons are based have generally employed common rather than comparable criteria of original learning for the recall and recognition groups (Postman, 1952).

#### EASY OR DIFFICULT RECOGNITION TESTS

In discussing recognition measures Deese (1958) pointed out that the recognition scores depend upon various artifacts of testing: "The probability of a correct selection on the part of the *S* is at the mercy of the particular choices he has to make [p. 240]." Similarly, Underwood (1949) pointed out:

If we required *S* to learn a list of adjectives and then place the adjectives among a group of nonsense syllables, *S* would probably show

very small loss in retention. Obviously the similarity of the test material to the other material is an important variable which determines the recognition score [p. 512].

Everyone who has given multiple-choice tests of recognition in academic courses knows that one can manipulate the difficulty level of an item simply by varying the incorrect alternatives. The effect of degree of differentiation among correct and incorrect alternatives upon the threshold of recognition and upon recognition retention curves has not been investigated systematically, but relevant data were reported by Lehmann (1888-89). He showed that net recognition scores for identification of gray colors declined from 87% correct to 17% correct as the wrong stimuli presented to *S* became more similar to the correct ones. Similarly, Seward (1928) showed that frequency, speed, and confidence of correct recognition responses depended upon the similarity of incorrect to correct items. Postman, Jenkins, and Postman (1948) showed that when one of the alternatives offered on a recognition test is more similar to the correct alternative, it receives a disproportionate share of the error choices. These data suggest that the number of trials needed to reach the threshold of a recognition test may be increased by increasing the similarity among the right and wrong alternatives of the test. Verbal material offers limited opportunities for manipulating the variable of similarity among responses along a continuum. Other symbolic material presents better opportunities for this and is therefore more suitable for an investigation of this variable. In a study comparing retention curves for recall and recognition measures (Bahrick & Bahrick, in press), differentiation among the correct and incorrect stimuli and responses in a paired-associates learning task was varied. The paired-associate method of pres-

entation was used during training. Stimulus material consisted of  $8 \times 8$  matrices in which 1 of the 64 cells was colored black, the remainder left blank. The responses to be matched with the stimuli were randomly chosen 3-digit numbers. One group of subjects was trained to the threshold of anticipation, that is, the first trial on which they correctly listed the 3-digit numbers paired with each of the matrices. The other two groups were trained, respectively, to an easy and to a difficult recognition threshold, as determined by two different recognition tests. In both the easy and the difficult recognition tests subjects were required to identify the correct matrix (stimulus) among five matrices presented in one row of the test blank, and to identify the associated number among five 3-digit numbers in the same row. In the easy recognition test the four wrong matrices differed from the correct one in that the stimulus position was displaced by a distance of three cells; in the difficult recognition test the displacement was only one cell removed. Similarly, the easy recognition test provided incorrect response numbers which shared none of the 3 digits with the correct number, whereas the alternative response numbers for the difficult test shared 2 out of 3 digits with the correct number. The recognition tests were administered during each 130-second intertrial interval. The same intertrial interval was used for the anticipation group, but this group took no recognition tests during the intervals. The mean number of trials required to reach the easy recognition threshold was 8.0, the anticipation threshold 9.3, the difficult recognition threshold 11.5. The group differences were significant ( $p < .01$ ). The data thus provided an instance in which a threshold of recognition lies above the threshold of anticipation for the same material.

In the Bahrack and Bahrack study

retention was measured after 2 hours, 2 days, or 2 weeks by the same method used during the training of each group. All three groups showed negatively accelerated retention curves, and the percentage correctly retained by the Anticipation group was lower than for the Easy Recognition group, but higher than for the Difficult Recognition group. This was true for all three time intervals.

This study clearly showed that thresholds for anticipation are not necessarily higher than for recognition, and the results suggest that differences in the slopes of retention curves are determined by overlearning with respect to the threshold level, rather than by the retention measurement per se. The following examination of artifacts produced by dichotomous scoring procedures makes this even more apparent.

#### CHANGING SENSITIVITY OF DICHOTOMOUS SCORES

It has been demonstrated that the slope of learning curves based upon dichotomous scores is a predictable artifact of the changing sensitivity of the indicants at different stages of training (Bahrack et al., 1957). The slopes of such curves therefore can not provide a valid index of relative amounts of learning at different stages of training. An analogous situation exists with respect to retention curves. Measures of anticipation, recall, and recognition are dichotomous indicants in that they tell us only which associations are above and which are below the threshold reflected by the response measure. A subject either recalls a nonsense syllable or he does not recall it. No further differentiation of associative strength is obtained with a recall score. The same is true for indicants of anticipation and recognition, though the particular threshold dividing correct from incorrect responses may be different for each of these indicants, and

in the case of recognition scores will depend upon the difficulty of the differentiation required by a particular test. The sensitivity of each of these measures to losses of associative strength changes in time. When material has been greatly overlearned with respect to a given measure, so that most of the associations are far above threshold strength, the measure is insensitive. This means that the strength of the associations may weaken considerably in time, but the slope of the curve based upon this measure will not show it because only a few of the originally, weakest associations will be weakened sufficiently to pass below the threshold. The slope of the curve does not directly reflect the *amount* by which associations are weakened in time. The slope of the curve simply reflects the number of associations passing below the threshold per unit of time. Assuming greatest frequency for associations which are originally of medium strength, the measure will become more sensitive as time passes and these associations are further weakened. Still later, when the bulk of the distribution has already passed below the threshold, the measure again loses sensitivity, and the curve flattens out. If the anticipation and recognition thresholds for a particular task are widely separated, the time periods of maximal sensitivity for the respective curves will differ greatly, and as a consequence the slopes of the respective curves during a given time period will be very different. This has been commonly observed and has led to the mistaken conclusion that recognition measures yield curves which differ, *per se*, from those obtained by means of free recall or anticipation measures. If threshold level and the degree of learning with respect to the threshold are comparable, however, as is the case for the data reported by Bahrick and Bahrick, the slopes of the resulting anticipation

and recognition curves are comparable also.

Retention curves based upon anticipation, recognition, or free recall reflect the changing sensitivity of dichotomous measures. No general conclusions regarding relative amounts of forgetting at different periods of time are permissible on the basis of such a curve. Such conclusions are valid only if they are based upon equal-interval scales. A dichotomous measure based upon a low threshold might show maximal forgetting over a time period during which the curve based upon a higher threshold has already approached an asymptote of zero.

A completely comparable analysis has been made for learning curves based upon time-on-target scores and other dichotomous indicants (Bahrick et al., 1957). The slopes of these curves could be predicted with good precision because evidence was available regarding empirical (near normal) error-amplitude distributions. The distribution of the strength of associations is not precisely known for any stage of forgetting, but it is probably not normal if learning is terminated on the first trial on which all associations are above threshold strength. Precise prediction of the slopes of retention curves based upon dichotomous measures will have to wait until empirical distributions of associative strengths reflecting inter- and intraindividual differences are obtained for various types of material and for different degrees of original learning.

It would follow from the preceding analysis, however, that the degree of homogeneity, as reflected by the combined inter- and intraindividual variance of the strength of various associations in a list, would partly determine the slope of retention curves based upon dichotomous indicants. The greater the variance, the fewer associations would pass through the threshold per unit of

time, and thus the less steep the slope of the retention curve. Conversely, if all associations in a list had exactly the same strength, that is, if the variance were zero, a step-function drop from 100% to 0% retention would be predicted for a curve of an individual subject regardless of the rate at which associations are actually weakened in time. The time of this drop would of course depend upon the height of the particular threshold.

The strength of various associations within a list does not necessarily weaken equally in time. Rather, inter- and intra-individual variance may change as a function of differential interference processes, and in this way bring about curves of different slope for material of varying length, similarity, and meaningfulness. The interpretation of these curves, however, must take into account the scoring artifacts which have been discussed. This analysis cannot be directly applied to curves based upon savings scores, the only nondichotomous indicants of retention.

REFERENCES

ACHILLES, E. M. Experimental studies in recall and recognition. *Arch. Psychol.*, 1920, 6, No. 44.

ANDREAS, B. G. *Experimental psychology*. New York: Wiley, 1960.

BAHRICK, H. P., & BAHRICK, P. O. A re-examination of interrelations among measures of retention. *Quart. J. exp. Psychol.*, in press.

BAHRICK, H. P., FITTS, P. M., & BRIGGS, G. E. Learning curves: Facts or artifacts? *Psychol. Bull.*, 1957, 54, 256-268.

BRITT, S. H. Retroactive inhibition: A review of the literature. *Psychol. Bull.*, 1935, 32, 381-440.

CLARKE, H. M. Recall and recognition for faces and names. *J. appl. Psychol.*, 1934, 18, 757-763.

DEESE, J. *The psychology of learning*. New York: McGraw-Hill, 1958.

EBBINGHAUS, H. *Memory: A contribution to experimental psychology*. (Trans. by H. A. Ruger & C. E. Bussenius) New York: Teachers College, Columbia University, 1913.

HOLLINGWORTH, H. L. Characteristic difference between recall and recognition. *Amer. J. Psychol.*, 1913, 24, 533-544.

LEHMANN, A. Ueber Wiedererkennen. [Concerning recognition.] *Phil. Stud. (Wundt)*, 1888-89, 5, 96-156.

LUH, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, 31(3, Whole No. 142).

MILER, ADRIENNE. Vergleich der Vergessenskurven für Reproduzieren und Wiedererkennen von sinnlosem Material. [Comparing retention curves for recall and recognition of meaningless material.] *Z. exp. angew. Psychol.*, 1957, 7, 29-38.

MYERS, G. C. A comparative study of recognition and recall. *Psychol. Rev.*, 1914, 21, 442-456.

POSTMAN, K. Retroactive inhibition in recall and recognition. *J. exp. Psychol.*, 1952, 44, 165-169.

POSTMAN, L., & RAU, L. Retention as a function of the method of measurement. *U. Calif. Publ. Psychol.*, 1957, 8, 217-270.

POSTMAN, L. W., JENKINS, O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *Amer. J. Psychol.*, 1948, 61, 511-520.

SEWARD, G. H. Recognition time as a measure of confidence. *Arch. Psychol.*, N.Y., 1928, No. 99.

STALNAKER, J. M. Recognition and recall in a vocabulary test. *J. gen. Psychol.*, 1935, 46, 463-464.

STRONG, E. K. The effect of time-interval upon recognition memory. *Psychol. Rev.*, 1913, 20, 339-372.

UNDERWOOD, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.

(Received September 7, 1962)

## NOTE ON MULTIVARIATE METHODS FOR PROFILE ANALYSIS<sup>1</sup>

JOHN E. OVERALL  
*Kansas State University*

Caution must be exercised in the use of the  $d$  statistic and related procedures for the evaluation of profile data. The Pythagorean distance model requires comparable scale units and zero correlations among profile elements. A general model which can be employed appropriately with correlated multivariate data is described. The  $d$  statistic defined as the square root of the sum of squared differences between scores on separate profile elements is shown to be a special case requiring stringent assumptions if the resulting value is to be interpreted as representing interprofile distance in Euclidean space.

In a recent issue of the *Bulletin*, Nunnally (1962) has attempted to place profile analysis within "a mathematical and statistical framework." While he demonstrated a very interesting relationship between distance measures of profile similarity (Osgood & Suci, 1952; Cronbach & Gleser, 1953) and results obtained from inverse factor analysis methods (Stephenson, 1953), his work lends sanction to a very specialized approach which has in the past been emphasized in the pages of this journal to the exclusion of more general approaches. The purpose of this note is to recommend an alternative, more general model which has not been given adequate treatment by other writers.

The distance between two points in multidimensional space can be defined according to the Pythagorean theorem as the square root of the sum of squared differences in projections upon uncorrelated (comparable unit) reference axes. The  $d$  statistic, and hence Nunnally's "raw score" factor analysis, is said to be based upon such a geometric model. The "distance" between two multivariate profiles is computed as the

square root of the sum of squared differences between corresponding scores in the two profiles. The conception of the  $d$  statistic as representing distance in multidimensional space thus implies that the profile elements are uncorrelated and are expressed in comparable units.

The necessity for equivalence of scale units is easily recognized. Hardly anyone would fail to object to adding squared inches to squared feet or squared centimeters in obtaining a single distance index. In what units would the resulting distances be represented? The major problem is not simply that the scale of measurement will be arbitrary and unlike that of any of the original measures; the problem is that the arbitrary units differ for different profile pairs depending upon the relative degree of difference between them on the several separate profile components. Moreover, we recognize that it is very difficult to justify the assumption that scale units are comparable in most areas of psychological measurement. A solution which has frequently been employed is to transform all scales to standard deviation units. While this approach may not provide a completely satisfactory solution, it does introduce some meaningful basis for assuming comparability.

An even more difficult problem arises

<sup>1</sup> This work was accomplished during a period in which the investigator's efforts were supported by a Research Career Development Award, National Institute of Mental Health, Bethesda, Maryland.

from heterogeneous intercorrelations among profile elements. Distance is defined as the square root of the sum of squared differences in projections on *uncorrelated* reference axes. If correlations between profile elements exist, then the distance between two profiles in multi-dimensional measurement space is *not* the square root of the sum of squared differences in scores on the correlated profile components. If differential or heterogeneous intercorrelations exist, a very distorted picture of true distances between individuals may result from the simple summing of squared differences on the separate measures.

Cronbach and Gleser (1953) have discussed the problem of correlations among profile elements. They indicate that the  $d$  statistic computed over correlated profile elements is equivalent to a  $d$  statistic computed from the underlying orthogonal factor variates, each weighted according to the portion of the total variance accounted for by that factor. A major difficulty is that we seldom know the extent to which underlying orthogonal components are reflected in the profile elements so that we do not know what is being weighted by how much. What is more, this uncertain weighting of orthogonal components is precisely like computing interprofile distances from measurements which, while uncorrelated, are not of comparable units. Thus, we do not seem to avoid difficulty in relating the  $d$  statistic to the geometric model when profile elements are not independent. If one wishes to weight factor dimensions differentially, the more precise approach would seem to involve computing the orthogonal components and then weighting them with full knowledge and freedom concerning choice of the weighting factors.<sup>2</sup>

<sup>2</sup> In later writings some of the investigators who previously recommended use of the  $d$  statistic or its equivalent  $Q$  type covariance

### *Generalized Multivariate Approach to Profile Relationships*

An approach which appears somewhat more adequate than the  $d$  statistic, or the inverse factor analysis, is based upon a generalization of the distance function approach to the study of group differences (Mahalanobis, 1936). Consider the multivariate distribution of  $p$  correlated measurements over a heterogeneous population of individuals. The squared distance between any two profiles can be expressed in terms of a single linear combination of the multiple correlated measurements in units of the population variance for that composite measure.

$$D^2 = \frac{(\sum a_i d_i)^2}{\sum \sum a_i a_j \sigma_{ij}^2}$$

where  $d_i$  is the difference between scores in the  $i$ th profile element for two individuals,  $a_i$  is the  $i$ th compounding coefficient, and  $\sigma_{ij}^2$  is the covariance between the  $i$ th and  $j$ th profile elements.

Taking partial derivative with respect to the  $a_i$  and setting equal to zero results in a set of simultaneous equations which can be solved to obtain values for the  $a_i$  which will maximize the squared distance between the profiles relative to the population variance. The solution is represented in the following matrix equation which can be solved for relative values of  $a_i$ .

$$a = C^{-1}d$$

where  $C^{-1}$  is the inverse of the variance-covariance matrix for the  $p$  profile elements, and  $d$  is the vector of difference scores.

The squared difference between profiles in the correlated measurement space is then

$$D^2 = a_1 d_1 + a_2 d_2 + \dots + a_p d_p = d' C^{-1} d.$$

have become concerned with problems inherent in the approach (Cronbach, 1958).

Obviously, if the profile elements are uncorrelated and have equal variance, the matrix  $C^{-1}$  will be proportional to an identity matrix; for example,  $C^{-1} = k \cdot I$ . In this special case,  $D^2$  is simply the sum of squared differences between the separate profile elements.

$$D^2 = d'C^{-1}d = d'd = \sum_{i=1}^p d_i^2.$$

This is the special case with which the  $d$  statistic and Nunnally's "raw score" factor analysis are concerned. When the conditions of equal variances and zero covariances are reasonably approached, the specialized solution may prove quite useful. On the other hand, it would appear important to be aware of conditions under which the less general approach may prove misleading.

#### *Vector Product Approach to Profile Relationships*

It has been considered that the squared distance between two points in multidimensional space can be represented as

$$D^2 = d'C^{-1}d.$$

Recognizing that the vector  $d$  is the difference between two score vectors, we can also represent the distance between the two profiles in terms of their respective distances from origin and the angle between the vectors.

$$\begin{aligned} D^2 &= (x_1 - x_2)'C^{-1}(x_1 - x_2) \\ &= x_1'C^{-1}x_1 + x_2'C^{-1}x_2 - 2x_1'C^{-1}x_2 \end{aligned}$$

where  $x_1'C^{-1}x_1 = L_1^2$  and  $x_2'C^{-1}x_2 = L_2^2$  represent the squared distances of the respective profiles from origin (i.e., the length of the vectors) and  $x_1'C^{-1}x_2 = \cosine \Theta \cdot L_1 \cdot L_2$  represents the cosine of angular separation times the product of the vector lengths.

Considering a profile to be a point in multidimensional space, we recognize

that it can be defined completely by direction and distance from origin. Thus, the three characteristics of "level," "variance," and "pattern" (Os-good & Suci, 1952) are replaced by the two factors of "direction" and "distance" from origin.

In some problem areas, and psychiatric classification may prove to be one, direction from origin may be the only important basis for judging profile similarity. Patients are of a similar type when their profiles tend to be proportional irrespective of distance from origin. It is possible to eliminate distance-from-origin as a variable in judging profile similarity by simply scaling all vectors to unit length. This is accomplished by dividing each profile element by the square root of  $x_i'C^{-1}x_i = L^2$ .

$$\begin{aligned} D^2 &= x_1'C^{-1}x_1 + x_2'C^{-1}x_2 - 2x_1'C^{-1}x_2 \\ &= 2(1 - x_1'C^{-1}x_2) \end{aligned}$$

where  $x_1'C^{-1}x_1 = x_2'C^{-1}x_2 = 1.0$ .

Normalizing the profile vectors is equivalent to projecting all profile points to the surface of a unit radius hypersphere. The distances between points are then solely a function of the angles between the vectors.

In the special case where profile elements have equal variances and zero covariances (e.g., the matrix  $C^{-1}$  is proportional to  $I$ ), the squared distance expressed in terms of vector lengths and cosines is

$$D^2 = \sum X_1^2 + \sum X_2^2 - 2\sum X_1X_2.$$

This is where the raw cross-product analysis proposed by Nunnally comes into the general picture.

Normalizing profile vectors by dividing each element by the square root of the raw sum of squares for all elements within the profile results in

$$D^2 = 2(1 - \sum X_1X_2),$$

where  $\sum X_1^2 = \sum X_2^2 = 1.0$ .

Thus, where it can be assumed that profile elements are uncorrelated and have equal variance, normalizing profile vectors results in  $D^2$  values which are a function of the raw cross-product term only.

### *Factor Analysis Approach to Profile Relationships*

Direct evaluation of  $D^2 = d'C^{-1}d$  may prove to be computationally impractical when the number of profiles, or the number of elements within profiles, is large. It is possible to transform any set of  $p$  correlated measurements into  $q \leq p$  mutually uncorrelated variates having unit variance (Rao, 1952, p. 345). Any method of orthogonal factor analysis can be employed to accomplish this purpose provided that the total variances of the profile elements are included in the matrix which is factored (Overall, 1962). For the uncorrelated, unit variance factor variates, the variance-covariance matrix will be an identity, and  $D^2$  can be obtained as the sum of squared differences in scores on the uncorrelated factor variates.

$$D^2 = d'C^{-1}d = d'd = d_1^2 + d_2^2 + \dots + d_r^2,$$

where  $d$  is the difference vector for uncorrelated, unit variance factor scores.

In some instances we may be concerned with the "psychological" distance between objects or persons which may be different from the value obtained by giving equal weight to the orthogonal factor variates. For example, we may conceive that the psychological judgment of distance is based upon differential perceptual weighting of various factor dimensions. A weighted distance function may be desired.

$$D^2 = a_1 d_1^2 + a_2 d_2^2 + \dots + a_r d_r^2$$

where  $D^2$  is the square of the perceived psychological distance between two objects, the  $d_i^2$  are squared distances along independent factor dimensions, and the  $a_i$  are perceptual weighting factors.

If we have global distance judgments and simple distances between a number of pairs of objects on uncorrelated factor dimensions, we can form a set of  $n$  equations in  $r$  unknown and solve for the  $r$  values of  $a_i$  which will minimize error in accounting for "psychological" distance in terms of the simple factor dimensions. This would seem to be a more precise solution to the problem of perceptual weighting of independent factor dimensions than is afforded by the  $d$  statistic, which weights independent factor dimensions according to the arbitrary degree of representation of the factor dimensions in the profile.

### REFERENCES

- CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri and L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958.
- CRONBACH, L. J., & GLESER, G. C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, 50, 456-473.
- MAHALANOBIS, P. C. On the generalized distance in statistics. *Proc. Nat. Inst. Sci., India*, 1936, 12, 49-58.
- NUNNALLY, J. The analysis of profile data. *Psychol. Bull.*, 1962, 59, 311-319.
- OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean differences and profile information. *Psychol. Bull.*, 1952, 49, 251-262.
- OVERALL, J. E. Orthogonal factors and uncorrelated factor scores. *Psychol. Rep.*, 1962, 10, 651-662.
- RAO, C. R. *Advanced statistical methods in biometric research*. New York: Wiley, 1952.
- STEPHENSON, W. *The study of behavior: Q-technique and its methodology*. Chicago: Univer. Chicago Press, 1953.

(Received September 11, 1962)

## COLOR AS AN INDEPENDENT VARIABLE IN PERCEPTUAL RESEARCH

M. CARR PAYNE, JR.

Georgia Institute of Technology

A summary of the literature in which color or its characteristics (hue, luminance, saturation) have been treated as an independent variable with dependent variables of apparent distance, apparent size, apparent weight, apparent temperature, and visual acuity. Luminance appears to be the major cue to apparent size, distance, or weight. Hue is the more important cue to apparent temperature. A chromatic figure seen against an achromatic background produces best acuity. Implications of this literature are discussed.

Historically, visual perception, particularly space perception, has been described in terms of stimulus cues. These have been treated as independent variables in experiments in which judgments of size, distance, or acuity were the dependent variables (Bartley, 1958; Dember, 1960; Gibson, 1950; Graham, 1951). Color perception typically has been treated as a dependent variable, although studies of the cues of aerial perspective or light and shade might be considered to include certain characteristics of color. Koffka (1935) has pointed out that "the different aspects of our visual world, size, shape, colour, orientation, and localization, are constituted in a thoroughgoing mutual interdependence [p. 265]. This suggests that color perception should also be investigated as a stimulus cue in order to determine its influence on perception. The present paper summarizes the literature in which color has been treated as an independent variable in regard to dependent variables of apparent distance, apparent size, apparent weight, apparent temperature, and visual acuity.

### EFFECTS ON APPARENT DISTANCE

Two colored stimuli adjacent to each other may appear to lie in different planes. Three explanations have been offered

for this phenomenon: (a) Chromatic aberration, in the refractive media of the eye radiation of short wavelengths (i.e., producing blue), focuses images within the eye closer to the lens than radiation of long wavelengths (i.e., producing red). The same sort of effect appears for distance, as radiation from distant targets projects images closer to the lens than does radiation from nearby targets. This forms a possible basis for learning to discriminate (Bartley, 1958, p. 165). Thus, a red object would appear to be closer than a blue object of the same physical size and physical distance. (b) *Eindringlichkeit*—Katz (1935) believed that colors vary in a quality which he called *Eindringlichkeit* (translated "insistency" after Titchener, 1916) and that this quality affects the apparent distance to which the color will be referred. Koffka (1935, p. 113) and Titchener (1916, p. 55) both pointed out that insistency varies with intensity as well as with other qualities (e.g., whiteness for Koffka, clearness conjoined with quality, or with intensity, or with quality and intensity together for Titchener). (c) More recently, a third explanation has been offered (Smith, W. M., as quoted in Ittelson, 1960, p. 101). This suggests that the more a color approximates the background (either in hue, luminance, saturation, or a com-

bination) the more it appears to recede towards the background. The more a color differs from the background, the farther it stands out from the background. Thus, of two colored stimulus objects, the one most like the background would appear to be more distant than the other.

Most studies in the literature (Belaiw-Exemplarsky, 1925; Johns & Sumner, 1948; Katz, 1935; Leventhal, 1953; Luckiesh, 1918; Mount, Case, Sander-son, & Brenner, 1956) reported objects of longer wavelengths to be perceived as nearer than objects of shorter wavelengths. Zajac (1953) reported the same finding when monocular distance judgments are made through a colored filter. Zajac took this to support the theory of chromatic aberration as a cue to depth.

Pillsbury and Schaeffer (1937) found results which contradict the usual finding. Using red and blue lights, equated in intensity as well as possible, and compensating size of the lights at different distances to give the same retinal image regardless of distance they found that 11 of 15 observers reported blue to be nearer than red although the red light was actually nearer. All of the observers mentioned luminance of the color that they judged to be nearer as the reason for that judgment. If the blue light actually was of higher luminance than the red light then Pillsbury and Schaeffer's results are understandable, as colors of higher luminance appear nearer than colors of lower luminance.

Early studies of color effects did not equate stimulus objects in intensity. Mount et al. (1956) overcame this objection. A standard stimulus, either a bright gray or a dark gray square, was placed 200 feet in front of the observer. Four hues (red, yellow, green, and blue) and four grays, chosen so that one of the grays matched one of the hues in luminance, were used as comparison stimuli. Each hue was perceived as

closer than its corresponding gray. The magnitude of this effect was 3 feet or 1.5% of the standard distance. All of the comparison stimuli, both gray and hued, appeared closer when compared with the darker standard stimulus than when compared with the lighter one. This contradicts results obtained by Edwards (1955) who reported failure to find consistent evidence that some colors are drawn nearer to the observer than others on the Howard-Dohman apparatus when they are compared with a gray standard. Mount and associates' study indicated that hue does have an effect upon apparent distance, but the experimental design did not permit an estimate of the difference in effect among hues.

Luminance appears to be a major cue to distance regardless of hue effects (Ames, 1946; Ashley, 1898; Pfeiffer, 1937). Vernon (1952, p. 105) reported that, in general, of two light boxes of different luminance, the higher luminance appeared nearer, although the ratio of apparent luminances varied at different distances of the two boxes. Taylor and Sumner (1945) placed colored papers on the movable pole of the Howard-Dohman apparatus. The observers equated it in distance with a gray stationary pole. All colors were seen behind the gray pole. Luminance seemed to provide the major cue, as a rank-order correlation of .99 was obtained between luminance measures of the colors and their average distances from the stationary pole. Johns and Sumner (1948) verified this result using the same apparatus but placing the colors on the stationary pole and the gray on the movable pole. Luminance correlated .66 with distances of colors from the observer's eyes. In these studies, colors of higher luminance appeared nearer than colors of lower luminance at a constant distance.

Coules (1955) investigated the relationship between luminance and apparent distance further by varying the

luminance ratios between a stationary standard stimulus and a movable variable stimulus. The variable stimulus at each distance presented a constant size to the observer's eyes. When the variable stimulus was of higher luminance than the standard stimulus, regardless of its distance from the standard, more judgments of the variable's appearing nearer were made under monocular conditions than under binocular. This suggests that the cue of luminance affects distance judgments independently of binocular cues. When the two stimuli were of the same luminance, the frequency of responses, indicating the variable stimulus to be nearer than the standard, did not vary as the variable was brought nearer. Under both monocular and binocular conditions, the percentage frequency of responses, of the variable stimulus' appearing nearer than the standard, increased as the logarithm of the ratio between the two luminances was increased. In a second experiment, Coules verified this relationship at three absolute luminance levels.

#### EFFECTS ON APPARENT SIZE

Warden and Flynn (1926) had observers rank eight cartons covered with colored Hering papers according to apparent size. They found no effects on apparent size due to color, although they reported an effect due to specific serial arrangement of the colors while the observer was ranking them. Gundlach and Macoubrey (1931) replicated this study using cartons covered with Milton Bradley papers. Regardless of arrangement, these authors found a significant effect of color on apparent size. To test this further they took paired-comparison data from the same observers on the same stimuli. Correlations of ranks from the first part of the experiment with those from the paired-comparison data was .93.

Bevan and Dukes (1953) tested color

effects by mounting rectangular stimulus cards of four different sizes and four different colors in a natural setting at distances of 18–52 feet from the observer. The observer selected a card the same size as the stimulus card from among 14 neutral gray reference cards. Sizes of red and yellow stimuli were overestimated and were different from size estimates of blue and green stimuli which showed no reliable difference between estimated size and objective size. No significant difference was found between size estimates of red and yellow stimuli nor size estimates between blue and green stimuli. Deviations of estimated size from actual size varied with objective size but the interaction between objective size and color was not significant. Unfortunately, the authors did not test for the interaction between distance and color in evaluating effects on size.

The studies described above and several other studies (Franklin, 1956; Kanicheva, 1938, 1939; Mitra & Datta, 1939; Sato, 1955) have reported color to have an effect on apparent size. To date, no one has yet identified the effect of each dimension of color (hue, luminance, or saturation); although studies by Gundlach and Macoubrey (1931), Sato (1955), and Wallis (1935) have shown that luminance is significantly correlated with apparent size, objects of higher luminance being overestimated in size. Holway and Boring (1940) verified this with noncolored stimuli. Robinson (1954) found the increase in size due to luminance to be of the order of 3–4% under either monocular or binocular conditions. Increasing the absolute level of luminance of two stimuli decreased overestimation of the stimulus of higher luminance.

Using two levels of luminance (25 and .025 foot-lamberts), Chalmers (1953) showed that variability of judgments was not increased in either monocular or binocular judgments as luminance was reduced. The author described the effect

of reduced luminance as a regression to monocular vision, as at the lower luminance level binocular judgments of size seemed to follow the visual angle when the standard stimulus (isosceles triangle) was presented at seven different distances from 10 to 120 feet. Visual angle of the standard at each of these distances was the same. Visual angle is the only cue for monocular perception of the size of objects beyond 10 feet when secondary cues are removed as they were in this study.

As is the case with apparent distance, luminance appears to be the major cue to apparent size. There are no studies in the literature of the effects on apparent size of varying the hue under constant luminance conditions.

#### EFFECTS ON APPARENT WEIGHT

In an early study, Bullough (1907) reported that objects of different colors appear to be of different weights. DeCamp (1917) using paired comparisons found a slight tendency to judge a red or black block to be heavier than a yellow or blue block of the same weight when each colored block was compared with a gray block of the same size. Order of apparent weight from heavier to lighter was red, white, orange, violet, green, purple, blue, and yellow-green. A set of order of merit data in which observers simply lifted the blocks and placed them in order of apparent weight resulted in the order: red, yellow, green and blue. Large individual and sex differences appeared in this study. Kimura (1950) found similar orders in apparent weight when an order of merit procedure was used and when a paired-comparison method was used. The order of apparent heaviness in his data was black, blue, red, violet, orange, green, yellow, and white.

Warden and Flynn (1926) found that effects of color on apparent weight were greater than effects on apparent size.

The order of apparent weight (from heavier to lighter) which appeared in their data was black, red, purple, gray, blue, green, yellow, and white.

Koch (1928) used Milton Bradley standard colors presented at opposite ends of a fulcrum. The observer was asked to balance them. He found no consistent order of apparent weight for his observers considered as a group. Each individual, however, seemed to show a consistent order of apparent weights as a function of color.

Monroe (1925) reported that apparent weight varied inversely with luminance. Taylor (1930) comparing "looking" and "hefting" procedures found this to be true, if colored objects identical in size and shape were observed, but did not find this to be true if the objects were hefted. Intersubject reliability was higher with the looking procedure than it was with the hefting procedure. Rankings obtained by an order of merit procedure were similar to those obtained by paired comparisons.

None of the effects reported in studies conducted before 1950 was evaluated statistically. For this reason, Payne (1958, 1961) carried out a series of studies to determine effects of color on apparent weight by a looking procedure. The first study employed paired comparisons. Median rank of apparent weight for each color was compared among observers by a coefficient of concordance which showed apparent weight as a function of color to be significant at the .01 level for three sizes of stimuli. Order of apparent weight was (heavier to lighter) blue, red, yellow, reddish-purple, turquoise, and green. Rank-order correlations between the median ranks of apparent weight for cues of different sizes showed the ranks of apparent weight to be highly correlated among stimulus sizes. A rank-difference correlation coefficient, computed between median ranks of apparent weight for each color and the rank of reflectance of that

color, showed that the higher the reflectance the lighter the stimulus appeared ( $p = -.94$ ). A second experiment, using the method of single stimuli, verified conclusions from the paired-comparison data. No significant relationship was shown between order of apparent weight and preference for colors.

The single stimuli procedure was repeated under controlled illumination conditions with 2-inch cubes covered with Munsell papers matched in luminance. Analysis of variance showed hues to be significant at the .01 level. However, yellow, green, and gray did not differ significantly from one another in their effect on apparent weight, nor did blue, red, or purple differ significantly from one another in their effect, although apparent weight of the latter group of hues did differ significantly from that of the former. This, hue does not seem to be as strong a cue to apparent weight as luminance.

Wright (1962) verified the effect of luminance using West German observers. He used an Osgood semantic differential technique. Each observer judged a single 3-inch square surface color on each of 24 adjective-pair rating forms. Forty-five chromatic colors, black, white, and gray, selected so as to be relatively uncorrelated with each other on the color dimensions of hue, luminance, and saturation, were rated in this manner. Use of German observers eliminated the confusion among English-speaking observers about the meaning of "lightness" (e.g., does it mean light-heavy or light-dark ratings?). Analysis of the partial regression of average *leicht/schwer* ratings on hue, luminance, and saturation showed a well-defined luminance effect and a less well-defined saturation and hue effect.

#### EFFECTS ON APPARENT TEMPERATURE

Writers in the early literature (Bulough, 1908; Lewinski, 1938; Newhall, 1941; Ross, 1938; Stefanescu-Grainger,

1911; Tinker, 1938) are in agreement that, as hue moves from red to blue, apparent warmth moves from warm to cold. Morgensen and English (1926) asked observers to judge the temperature of heating coils wrapped in paper of different colors. The coils were in fact of equal temperature. Observers judged green and blue coils to be less cool than purple coils. Red, orange, and yellow fell between the two extremes. No statistical tests were presented.

Kimura (1950) using both an order of merit and a paired-comparison method found the decreasing order of apparent warmth to be: red, orange, yellow, violet, green, black, blue, and white.

Berry (1960) conducted an experiment in which the observer was led to believe that effects of colored illumination upon tracking performance was being studied. The observer, who drove an American Automobile Association auto trainer, was instructed to push a switch when the temperature in the room rose to a point at which he felt uncomfortably warm. Temperature was gradually changed, but the changes were the same for each of five colors of diffuse illumination: green, blue, yellow, amber, and white. Luminance of each illuminant was the same. There were no significant differences in the Discomfort Index (based on wet- and dry-bulb temperatures at the moment when the observer indicated the onset of discomfort) between the colors of illumination. At the close of the test, observers were asked to rank the five colors according to the amount of heat they transmitted. These rankings were significant ( $p < .01$ ). Green and blue were ranked almost identically as the coolest, white was at an intermediate position, and yellow and amber were ranked almost identically as warmer. Thus, observers persisted in the conventional belief that green and blue are "cool" colors when asked to rank colors, although they did not experience them any differently from other

colors in the levels of heat the observer would tolerate.

Wright (1962) analyzed his semantic differential data for the polar adjectives, *kalt/warm*. He found that hue, luminance, and saturation accounted for at least half of the consistent variation while for apparent weight they account for about two thirds. There is a well-defined hue effect on apparent warmth independent of luminance or saturation. As reported in the early literature, as hue was moved from red to blue apparent warmth moved from *warm* to *kalt*. Effects of luminance and saturation were less definitive than those for hue. However, increasing warmth ratings did correspond with increasing saturation. In judgments of apparent temperature hue appears to be the major cue.

#### EFFECTS ON VISUAL ACUITY

The effects of stimulus color on acuity may vary according to the relations between the test object (stimulus) and its surround or background. The present paper will describe only studies in which a stationary stimulus figure seen against a flat background was employed.

If colored illumination is used, both the test object and its surround are affected. Under these conditions Koffka and Harrower (1932) found acuity values to be higher for targets seen as red, white, or yellow than they were for targets seen as blue or black. Ferree and Rand (1931, 1932), using white broken circle test objects, found acuity to vary with the color at both low (.075 foot-candle) and higher (.3116 foot-candle) illumination levels. With luminance kept constant the order of merit of acuity for hues (in millimicrons) was: red (666), orange (624), green (522), blue-green (500), and blue (488) for the lower illumination. At the higher illumination value the order was yellow, yellow-green, orange, green, and red. Blue and blue-green were not tested at

this intensity. These orders of colors were the same whether saturation varied or was constant among the colors. However, differences in acuity for different hues were less when equal saturations were used. Jonnard, Faillie, and Vial (1937), using Landolt rings, verified the finding that yellow illumination as compared with white illumination in different contrasts improved acuity of a light object on a dark ground. However, they found that it diminished acuity with a dark test object on a light ground.

When the amount of illumination is varied, visual acuity increases monotonically with illumination provided the illumination affects both the test object and its background so as to keep their luminance contrast constant (Konig, 1897; Shlaer, 1938). Weinberg and Lapinskaya (1941) have established an equation which permits one to calculate visual acuity as a function of illumination for monochromatic lights and their mixtures while taking account of the Purkinje phenomenon. The authors reported the effects of the Purkinje phenomenon on acuity to be much greater than its effects on luminance.

Acuity is better if the stimulus or test object is colored and is seen against an achromatic background than it is with achromatic stimuli and backgrounds. Schwarz (1951) found acuity for Landolt rings seen against a dark background to be best in the central region of the visual spectrum (i.e., yellows). Berbert (1958) verified this using two self-luminous dots seen against a dark field to test acuity. The dots were centered at 450, 500, 550, 600, or 650 millimicrons in hue. Observers were dark adapted before the experiment. Each hue was tested through the whole available range of luminances (3-decibel steps) during each session. Only one hue was used in one session. Best acuity, that is, lowest angular subtense of gap between dots when they appeared to be just touching, occurred at slightly dif-

ferent luminance values for each hue. However, the best acuity occurred at about the same increment above its corresponding operational threshold for each hue. The midregion of the visible spectrum (500–600 millimicrons) produced better acuity than did the extremes. In each observer the best acuity measured in the midregion of the spectrum was superior to acuity measured with an achromatic stimulus. A supplementary experiment in which measurements were made with all colors and luminances at each session confirmed these conclusions.

MacAdam (Stiles, 1947) reported that in Landolt rings, in which acuity measurements depend on discrimination in small areas, color contrast between a ring of one color and a background of another has much less effect than luminance contrast. Using colored rings mounted on an equiluminous gray background he (MacAdam, 1949) found colors of rings—expressed on the CIE (Commission Internationale d'Éclairage) diagram—whose orientations were reported correctly as frequently as those of equally large gray rings that were lighter than the same gray background. Using these figures as measures of chromatic contrast MacAdam concluded that:

When the target and background are not equiluminous, the acuity is the same as that produced by an achromatic contrast equal to the square root of the sum of squares of the luminous contrast and the chromatic contrast [p. 172].

Luminance has been shown to affect visual acuity (Berbert, 1958; Berger, 1941, 1942; König, 1897; Wilcox, 1936), but the shape of the mathematical function which describes this relationship has been questioned. Wilcox (1936) reported that, for a self-illuminated stimulus (parallel bars) seen against a dark background, increasing the illumination of the stimulus first lowered and then raised acuity with a dis-

tinct minimum value at medium intensities, that is, "best acuity" at these intensities. With a dark stimulus seen against an illuminated background, increasing the intensity of the background resulted in a continuously increasing acuity, but the rate of increase became progressively less as intensity increased. Wilcox interpreted his results in terms of irradiation effects which alter the apparent area of the parallel bars at higher illuminations. Since the minimum separation between the bars at which a gap was seen was used as a measure of acuity in this study, a change in apparent width of the bars would change the estimate of acuity.

Berger (1942) found results with self-illuminated Landolt rings or with points observed on an almost black background, which contradicted those of Wilcox. In both points and broken circle, the minimum visual angle decreased continuously with illumination, rather than increased with higher illuminations as occurred in Wilcox' study. This was true with either right or left eye or with both eyes. There were large individual differences in the absolute values obtained from the various subjects. Berger found smaller values with points as stimuli than he did with Landolt rings. Thus, he felt that most conventional measures of visual acuity should not be considered to measure "minimum resolving power of the eye," but that acuity and minimum resolving power should be considered apart from one another. He attributed the difference in his findings from those of Wilcox to be due to glare at the higher illumination levels in Wilcox' study.

Berbert (1958) attempted to resolve the contradictions between the results of Wilcox and Berger. Using bright dots on a dark field as stimuli he found visual acuity first rose to a maximum with increasing test object luminance and then fell off. Maximum acuity for the dots occurred near 3.2 millilamberts or at 10 times the luminance that maximum

acuity occurred in Wilcox' study. The difference in luminance, for maximum acuity when points instead of bars are used, is probably due to the considerable difference in subtense of these test objects.

### DISCUSSION

The major cue to apparent size, distance, and weight appears to be luminance. Lighter stimulus objects appear to be larger, nearer, and lighter in weight than darker stimulus objects. This characteristic appears to function regardless of hue of the stimulus object. On the other hand, hue is the more important cue to apparent temperature. Best acuity seems to occur with a chromatic figure seen against an achromatic background. There are few data available on the interactions between hue and luminance.

The finding that luminance is the more important cue to apparent size, distance, and weight leads us to seek a common functional relationship for these three dependent variables. Interaction between apparent size and apparent distance are well known (see Osgood, 1953, pp. 261-271). The fact that lighter objects appear to be nearer as well as larger than darker objects of the same physical size seen at the same distance is simply an example of this interaction. In a similar manner, objects of higher luminance appear both larger and lighter than objects of the same size and shape which are of lower luminance. Experiments on the size-weight illusion (Crutchfield, Woodworth, & Albrecht, 1955) have shown that larger objects of the same objective weight as smaller objects seem to be substantially lighter in weight than the smaller objects. Thus, we would expect influence of luminance on apparent weight from the finding of influence of luminance on apparent size and the size-weight illusion.

Rubin (1921) introduced the concept

of figure-ground into the study of perception. It might be argued that any characteristic which made a stimulus object stand out from its background, that is, made it more "figurelike," would make that stimulus object appear nearer than another stimulus object which differed from it only in that characteristic. Katz (1935) would explain the variation of apparent distance (and thus of apparent size and weight) with luminance on this basis by pointing out that *Eindringlichkeit* varies with intensity. *Eindringlichkeit* is considered to be an identifiable stimulus quality. Thus, an object of greater luminance would have greater *Eindringlichkeit* than one of lesser luminance and would appear nearer.

Gibson and Waddell (1952) have pointed out that an image becomes differentiated by means of "gradients of luminous intensity" which serve as transitions between relatively dark and relatively light regions of the visual field. A fully differentiated image is one in which the gradients of intensity are as steep as possible. Under these conditions the image would appear as a sharply defined figure against a background. At the other extreme, if the gradients of luminous intensity are zero, then the phenomenal experience is that of a uniform or homogeneous visual field or a fully undifferentiated image. Under these conditions there would be no figure at all. Between these extremes the gradient of intensity should determine how clearly a stimulus is perceived as a figure against a background. The less the intensity difference between figure and ground the less the figure would be differentiated and the more likely it is to be seen at the same distance as the background. The greater the difference between the luminance of two figures seen against a common background the farther apart the two figures should appear, the brighter one appearing closer than the darker one. Smith

(Ittelson, 1960, p. 101) supported this argument and generalized it to apply to gradients of hue and saturation as well as luminance.

Future research could well examine the interactions among gradients of hue, luminance, and saturation. When we can identify the extent to which these stimulus dimensions function together as perceptual cues we shall have a sounder basis for trying to discover *how* they came to serve as cues.

## REFERENCES

- AMES, A. Some demonstrations concerned with the origin and nature of our sensations. *Laboratory Manual*, 1946, Dartmouth Eye Institute, Hanover, New Hampshire.
- ASHLEY, M. L. Concerning the significance of intensity of light in visual estimates of depth. *Psychol. Rev.*, 1898, 5, 595-616.
- BARTLEY, S. H. *Principles of perception*. New York: Harper, 1958.
- BELAIEV-EXEMPLARSKY, S. Über die sogenannten "hervortretenden" Farben. *Z. Psychol.*, 1925, 96, 400-429.
- BERBERT, J. H. Visual acuity as a function of intensity for different hues. *Naval Res. Lab. Rep.*, 1958(May 26), No. 5104.
- BERGER, C. The dependency of visual acuity on illumination and its relation to the size and function of the retinal units. *Amer. J. Psychol.*, 1941, 54, 336-352.
- BERGER, C. A comparison under different degrees of brightness, of minimum visual distances of two luminous points and of a broken circle. *Amer. J. Psychol.*, 1942, 55, 354-370.
- BERRY, P. C. Effect of colored illumination upon perceived temperature. *Psychol. Res. Ass. Rep.*, 1960(May), No. 60-17.
- BEVAN, W., & DUKES, W. F. Color as a variable in the judgment of size. *Amer. J. Psychol.*, 1953, 66, 283-288.
- BULLOUGH, E. On the apparent heaviness of colours, a contribution to the esthetics of colour. *Brit. J. Psychol.*, 1907, 2, 111-152.
- BULLOUGH, E. The perceptive problem in the aesthetic appreciation of single colors. *Brit. J. Psychol.*, 1908, 2, 406-463.
- CHALMERS, E. J. The role of brightness in primary size-distance perception. *Amer. J. Psychol.*, 1953, 66, 584-593.
- COULES, J. Effect of photometric brightness on judgments of distance. *J. exp. Psychol.*, 1955, 50, 19-26.
- CRUTCHFIELD, R. S., WOODWORTH, D. G., & ALBRECHT, R. A. Perceptual performance and the effective person. *USAF PTRC tech. Note*, 1955.
- DECAMP, J. E. The influence of color on apparent weight: A preliminary study. *J. exp. Psychol.*, 1917, 2, 347-370.
- DEMBER, W. N. *The psychology of perception*. New York: Holt, 1960.
- EDWARDS, A. S. Effect of color on visual depth perception. *J. gen. Psychol.*, 1955, 52, 331-333.
- FERREE, C. E., & RAND, G. Visibility of objects as affected by color and composition of light: I. With lights of equal luminosity or brightness. *Personnel J.*, 1931, 9, 475-492.
- FERREE, C. E., & RAND, G. Visibility of objects as affected by color and composition of light: II. With lights equalized in both brightness and saturation. *Personnel J.*, 1932, 10, 108-124.
- FRANKLIN, J. C. Effects of color on apparent size. *Amer. Psychologist*, 1956, 11, 398. (Abstract)
- GIBSON, J. J. *The perception of the visual world*. Boston, Mass.: Houghton Mifflin, 1950.
- GIBSON, J. J., & WADDELL, D. Homogeneous retinal stimulation and visual perception. *Amer. J. Psychol.*, 1952, 65, 263-271.
- GRAHAM, C. H. Visual perception. In S. S. Stevens (Ed.), *A handbook of experimental psychology*. New York: Wiley, 1951. Pp. 868-921.
- GUNDLACH, C., & MACOUBREY, C. The effect of color on apparent size. *Amer. J. Psychol.*, 1931, 63, 109-111.
- HOLWAY, A. H., & BORING, E. G. The dependence of apparent visual size upon illumination. *Amer. J. Psychol.*, 1940, 53, 587-589.
- ITTELSON, W. H. *Visual space perception*. New York: Springer, 1960.
- JOHNS, E. H., & SUMNER, F. C. Relation of the brightness differences of colors to their apparent distances. *J. Psychol.*, 1948, 26, 25-29.
- JONNARD, R., FAILLIE, R., & VIAL, H. Mesure de l'acuité visuelle en fonction de quelques contrastes en lumière colorée. *Travail hum.*, 1937, 5, 306-317.
- KANICHEVA, R. A. [The influence of color on size perception.] *Bekht. Inst. Brain Res., 20th Anniv.*, 1938, 63-64. (*Psychol. Abstr.*, 13:3499)
- KANICHEVA, R. A. [The influence of color on size perception.] *Trud. Inst. Mozga. Bekht.*, 1939, 9, 61-75. (*Psychol. Abstr.*, 13:3500)
- KATZ, D. *The world of colour*. London: Kegan Paul, 1935.
- KIMURA, T. Apparent warmth and heaviness of colours. *Jap. J. Psychol.*, 1950, 20, 33-36. (*Psychol. Abstr.*, 26:3782)

- KOCH, B. C. *The apparent weight of colors*. New York: Methodist Books, 1928.
- KOFFKA, K., & HARROWER, M. Color and organization. *Smith Coll. Stud. Psychol.*, 1932, No. 3.
- KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
- KONIG, A. Die Abhängigkeit der Sehschärfe von der Beleuchtungsintensität. *SB Preuss. Akad. Wiss.*, 1897, 32, 559-575.
- LEVENTHAL, M. Color as a variable in perceptual response: A study of depth perception and word recognition in schizophrenic and other psychiatric patients. Unpublished doctoral dissertation, University of Kansas, 1953.
- LEWINSKI, R. J. An investigation of individual responses to chromatic illumination. *J. Psychol.*, 1938, 6, 155-160.
- LUCKIESH, M. On "retiring" and "advancing" colors. *Amer. J. Psychol.*, 1918, 29, 182-186.
- MACADAM, D. L. Color discrimination and the influence of color contrast on visual acuity. *Rev. Opt., theor. instrum.*, 1949, 28, 161-173.
- MITRA, S. C., & DATTA, A. The influence of color on the estimation of area. *Indian J. Psychol.*, 1939, 14, 91-94.
- MONROE, M. The apparent weight of color and correlated phenomena. *Amer. J. Psychol.*, 1925, 36, 192-206.
- MORGENSEN, M. F., & ENGLISH, H. B. The apparent warmth of colors. *Amer. J. Psychol.*, 1926, 37, 427-428.
- MOUNT, G. E., CASE, H. W., SANDERSON, J. W., & BRENNER, R. Distance judgment of colored objects. *J. gen. Psychol.*, 1956, 55, 207-214.
- NEWHALL, S. C. Warmth and coolness of colors. *Psychol. Rec.*, 1941, 4, 198-212.
- OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford, 1953.
- PAYNE, M. C., JR. Apparent weight as a function of color. *Amer. J. Psychol.*, 1958, 71, 725-730.
- PAYNE, M. C., JR. Apparent weight as a function of hue. *Amer. J. Psychol.*, 1961, 74, 104-105.
- PFEIFFER, E. Über den Einfluss von Helligkeit und Farbigkeit auf die Streckenteilung. [The influence of brightness and color on distance judgment.] *Arch. ges. Psychol.*, 1937, 98, 51-72.
- PILLSBURY, W. B., & SCHAEFFER, B. R. A note on "advancing and retreating" colors. *Amer. J. Psychol.*, 1937, 49, 126-130.
- ROBINSON, E. J. The influence of photometric brightness on judgments of size. *Amer. J. Psychol.*, 1954, 67, 464-475.
- ROSS, R. T. Studies in the psychology of the theatre. *Psychol. Rec.*, 1938, 2, 127-190.
- RUBIN, E. *Visuell Wahrgenommene Figuren: Studien in psychologischer Analyse*. Copenhagen, Denmark: Gyldendal, 1921.
- SATO, R. The effect of color on the perception of size. *Tohoku psychol. Folia*, 1955, 14, 115-129. (*Psychol. Abstr.*, 30:375)
- SCHWARZ, F. Über die Messung der Sehschärfe und Prüfung der Sehleistung mit farbigen Sehproben. *Pflüg. Arch. ges. Physiol. Menschen Tiere*, 1951-52, 254, 144-154.
- SJØLAER, S. The relation between visual acuity and illumination. *J. gen. Physiol.*, 1938, 21, 165-188.
- STEFANESCU-GRAINGER, F. Experimentelle Untersuchungen zur Gefuehlsbetonung der Farben. *Psychol. Stud.*, 1911, 7, 284-335.
- STILES, W. S. Mechanisms of colour vision. *Nature*, 1947, 160, 664-666.
- TAYLOR, C. D. Visual perception versus visual plus kinaesthetic perception in judging colored weights. *J. gen. Psychol.*, 1930, 4, 229-246.
- TAYLOR, I. L., & SUMNER, F. C. Actual brightness and distance of individual colors when their apparent distance is held constant. *J. Psychol.*, 1945, 19, 79-85.
- TINKER, M. A. Effect of stimulus-texture upon apparent warmth and affective value of colors. *Amer. J. Psychol.*, 1938, 51, 532-535.
- TITCHENER, E. B. *A text-book of psychology*. New York: Macmillan, 1916.
- VERNON, M. D. *A further study of visual perception*. Cambridge: Cambridge Univer. Press, 1952.
- WALLIS, W. A. The influence of color on apparent size. *J. gen. Psychol.*, 1935, 13, 193-199.
- WARDEN, C. J., & FLYNN, E. L. The effect of color on apparent size and weight. *Amer. J. Psychol.*, 1926, 37, 398-401.
- WEINBERG, V. B., & LAPINSKAYA, E. A. [A study of visual acuity in feeble chromatic illuminations.] *Probl. Fiziol. Opt.*, 1941, 1, 169-171. (*Psychol. Abstr.*, 22:5292)
- WILCOX, W. W. An interpretation of the relation between visual acuity and light intensity. *J. gen. Psychol.*, 1936, 15, 405-435.
- WRIGHT, B. The influence of hue, lightness, and saturation on apparent warmth and weight. *Amer. J. Psychol.*, 1962, 75, 232-242.
- ZAJAC, J. L. Investigations on depth perception in indirect vision. *Brit. J. Psychol.*, 1953, 44, 132-144.

(Received September 22, 1962)

## PSYCHOLOGICAL CORRELATES OF BLOOD PRESSURE<sup>1</sup>

NOEL F. MCGINN

*Instituto Tecnológico y de Estudios Superiores de Occidente,  
Guadalajara, Mexico*

ERNEST HARBURG

*University of Michigan*

STEVO JULIUS

*University of Zagreb, Yugoslavia*

AND JACK M. MCLEOD

*University of Wisconsin*

As a physiological measure of emotion, blood pressure has been relatively neglected in recent years. This review considers 3 questions: Do certain stimuli produce replicable changes in blood pressure? Do cognitive states influence blood pressure reactivity? Is there a causative relationship between personality and high blood pressure? Available research indicates that the 1st 2 questions can be answered affirmatively, though there are a number of methodological problems yet to be resolved. No direct evidence is available for the 3rd question and tangential studies offer conflicting results.

The use of measures of blood pressure, as dependent and intervening variables in psychological research, has a long and exciting history dating back before the turn of the century. But psychologists do not seem to have used the measure of an individual's blood pressure to represent "emotional" changes as much as they have used other measures, for example, the galvanic skin response. One aim of this review is to encourage a fresh look at the meaning, measures, and results of studies of blood pressure.

The design of this review can be summarized in the following way. The first studies to be considered are pertinent to this question: Are there replicable changes in blood pressure when subjects are presented with certain classes of external stimuli? Secondly, are there char-

acteristic ways of responding to a given external stimulus as a function of certain kinds of cognitive states or processes? The review concludes by presenting studies which attempt to explore the major question: Do various psychological states have possible causative relationships with levels of sustained high blood pressure?

Perhaps the first review of the literature relating psychological and physiological variables was that by Angell and Thompson (1899) which together with that of Robbins (1919) discussed various psychological variables as they related to subjects' reactions on a plethysmograph. The most comprehensive review was done in 1929 by Darrow. Recent reviews, which consider limited portions of the literature relating psychological and physiological variables, are those by Moses, Daniels, and Nickerson (1956) and Shapiro (1960). Also available is an annotated bibliography of blood pressure research from 1920 to

<sup>1</sup> The review was done in conjunction with the senior author's doctoral dissertation at the University of Michigan. Appreciation is expressed to R. L. Isaacson of the University of Michigan who reviewed an earlier version of this paper.

1950 (Koller & Katz, 1952). Because hypertension is more frequently found in Negroes than in Caucasians, there is a separate literature which has been reviewed by Schulze and Schwab (1936) and Lennard and Glock (1957). The present attempt is to provide a comprehensive review of that literature which is specifically relevant to persons interested in relationships between psychological and blood pressure variables.

#### EMOTION-PRODUCING STIMULI AND BLOOD PRESSURE RESPONSES

Techniques for nonsurgical measurement of human arterial blood pressure were first described in 1896 by Riva Rocci (according to Bard, 1956). Given an instrument for easy and painless measurement, a variety of pressors was ingeniously conceived to produce changes in blood pressure. By "pressor" is meant any externally or internally applied stimulus expected to induce a blood pressure response. By way of introduction to the vast amount of literature in this general area, it would be useful to illustrate the sensitivity of blood pressure to various stimuli.

The first published research on blood pressure in American psychology is that by Marston (1917), a Massachusetts lawyer, who wanted to demonstrate the validity of measures of blood pressure change for detecting when a person was lying. His subjects were asked to tell either the "truth" or a "lie," embellished around one or two facts supplied by the experimenter. Subjects' introspections were used to show that blood pressure increased when the subjects began to attempt to deceive and decreased when they told the truth. Actually, it did not matter what the objective facts were; so long as the subjects thought they were deceiving or telling the truth their blood pressures were affected. Langfeld repeated the

experiment in 1920 and obtained the same results.

Even dreams and music have been associated with changes in blood pressure. MacWilliams (1923) found that blood pressure might rise while subjects slept and that this elevation was associated with subjects reporting (on being awakened) that they had been dreaming. Lovell (1941) showed that blood pressure went up following a loud, unexpected sound. Steinman, Jaggi, and Widmer (1955) played subjects recordings of Mozart and Beethoven and produced a decline in blood pressure. When the subjects listened to jazz and Japanese music, blood pressure went up. The literature abounds with results which indicate that blood pressure changes may be associated with many forms of psychological stimuli or stresses.

#### *Psychological Stress and Temporary Elevation of Blood Pressure*

A number of researches have shown fluctuation of blood pressure during various kinds of interviews. One of the earliest studies of this kind in American research is that by O'Hare (1920) in which he reported that his patients reacted to the discussion of relevant life problems with an elevation in blood pressure. Alexander (1939b) noticed that, when his patients had particularly disturbing sessions in analysis, their blood pressures were markedly elevated. Wolf, Pfeiffer, Ripley, Winter, and Wolff (1948) assert that, for their patients, blood pressure went up when unpleasant life experiences were discussed and went down with "nice" conversation. Pfeiffer and Wolff (1950) demonstrated rises in blood pressure when subjects discussed personally important topics with "threatening" significance.

Perhaps the most thorough study employing the psychiatric interview as a stimulus is that by Innes, Millar, and

Valentine (1959). The experimenters used simultaneous recording of the interview and physiological measurements to permit a matching of blood pressure changes with specific content areas in the interview. Contrary to the findings of other studies to be reviewed later, the investigators found no correlations between specific kinds of interview content and blood pressure changes. Blood pressure did fluctuate according to interview content, but the authors were not able to establish a coding category which would fit with the blood pressure changes. There was, however, a significant relationship between the use of self-referent pronouns (I, my, etc.) and increases in blood pressure, regardless of the content of the sentence in which the pronoun appeared or the context of the discussion. Innes, Millar, and Valentine's subjects were both normals and neurotics, and it is possible that such a mixture obscured differences between the two groups. Their failure to create a code which would work may be viewed as a criticism of much of the work using the interview technique.

There are some more recent studies which take advantage of objective events posing threats to the physical or psychological life of people who experience them. For example, Graham (1945) measured the blood pressure of soldiers who had seen long combat in the African desert campaign in World War II and found 27% showed high diastolic blood pressure (>100 millimeters) for several months afterward. Ruskin, Beard, and Schaffer (1948) found that the diastolic blood pressure of 56% of Texas City survivors was markedly elevated (>95 millimeters) for about 1-2 weeks after the tremendous explosion there. A rather unusual experimental environment was created as a stimulus for parachutists by Grimak (1959). He hypnotized these

men and had them relive a jump experience. Continuous blood pressure readings were taken. Blood pressures went up when the plane took off, stayed high during the flight, went down when the subjects jumped and felt their chute open. Readings after subjects hit the ground were still higher than those taken before takeoff.

#### *Specific Blood Pressure Responses to Emotion-Producing Stimuli*

Cannon (1932) suggested that emotion was a generic term to apply to an emergency pattern of physiological changes regardless of the psychological components, for example, anger or fear. Ax (1953) attempted to explore physiological differences accompanying stimuli specifically designed to arouse anger and fear separately. The anger stimulus consisted of insulting behavior to the subject by a laboratory technician who alternately jostled, verbally insulted, and was generally nasty.

The fear stimulus consisted of a gradual intermittent shock stimulus to the little finger which never reached an intensity sufficient to cause pain. When the subject reported the sensation, the experimenter expressed surprise, checked the equipment, pressed a key which caused sparks to jump near the subjects, then exclaimed with alarm that this was dangerous high voltage short circuit [Ax, 1953, p. 35].

The same subjects were used in both the anger and fear situations. Fourteen physiological measures were taken, and frequency of changes, rather than amplitude, was used to measure response. The number of diastolic blood pressure, heart rate, galvanic skin response, and muscle tension increases was greater for anger than for the fear condition. Similarly Schachter (1957) exposed subjects to situations intended to produce emotions of pain, fear, and anger in that order. For the pain experience, subjects kept their hands in a bucket of ice water (3 degrees centigrade for 1 minute). The

anger and fear situations were similar to those used by Ax. Subjects responded to these stimulus conditions with a significant increase in diastolic blood pressure to the pain and anger conditions and a significant systolic increase to the fear condition. Another experiment by Funkenstein, King, and Drollette (1957) used frustrating tasks to induce anger. Results were similar to Schachter's.

Buss (1961) compares the results of the above three studies in terms of blood pressure changes. He asserts that in general these experiments support the notion that physiological emotion apparently is not as Cannon hypothesized, that is, all the same emergency pattern. The results also indicate that for fear (or anxiety) stimuli, at least, there may be a consistent blood pressure pattern of response: a rise in systolic pressure (and also heart rate) and a smaller increase in diastolic pressure. The results describing the blood pressure patterns accompanying anger are less consistent in these three studies.

#### PSYCHOLOGICAL FACTORS AS INTERVENING VARIABLES

In this section studies shall be reviewed in which psychological variables are used as intervening variables in the research. A frequently used clinical technique for assessing the dependent variable, blood pressure reactivity, is the cold pressor test. The subject immerses his hand (or foot) to the wrist in crushed ice and water at a temperature of 4-6 degrees centigrade for at least 1 minute. Subjects experience various degrees of pain, but with no ill effects. All subjects show some increase in blood pressure. While reaction time and rate of recovery may be thought by many psychologists to be important variables with this physical pressor, most research with the cold

pressor test has used amount of change from resting level as the significant variable. This is called "reactivity" as contrasted to the absolute blood pressure level reached.

There are several studies which seem to show that reactivity to the cold pressor test varies according to the psychological characteristics of the subjects. For example, White and Gildea (1937) observed that patients with somatic symptoms of anxiety (e.g., finger twitching) showed a greater reaction to the cold pressor test than did patients without these symptoms. Malmö and Shagass (1952) compared the reactivity of various types of psychiatric patients. Subjects diagnosed as neurotic showed greater increases of blood pressure on the cold pressor than did psychotics or a control group of normals and took longer to recover. More recently Harburg (1962) observed a significant association between college males who reported higher hostility in minor interpersonal conflicts and higher diastolic rise on the cold pressor test.

Cranston, Chalmers, Taylor, Henschel, and Keys (1949) combined the cold pressor test and a stressful interview. Subjects were given six to eight cold pressor tests on different occasions to establish their usual degree of reactivity. Subjects then experienced an interview in which they talked about topics which the investigators judged to be emotionally relevant for them. The interview is reported to have evoked hostility and resentment in most subjects. Immediately after the interview, the subjects again took the cold pressor test. Blood pressure levels were significantly higher. The process by which an emotion such as hostility or resentment heightens the effects of pain is not clearly understood. Also, it might be that the cold pressor was merely the final

stress necessary to produce an elevation of blood pressure.

Other kinds of physical stress conditions are employed to produce blood pressure reactions. For example, Stevenson, Duncan, Flynn, and Wolf (1952) had subjects perform a standardized exercise, taking blood pressures before and after. Subjects who were rated by psychiatrists as "relaxed" showed less increase in blood pressure than those rated as "preoccupied" or already in therapy for anxiety symptoms. Thurrell, Greenfield, and Roessler (1961) rated psychiatric patients (none of whom were on drug therapy) on degree of "psychological responsivity" based on a 15-minute interview. Subjects rated as "hyperresponsive" showed a greater variation of blood pressure from a base line following an injection of mecholyl than did those rated "hypo-responsive."

#### ARE THERE PSYCHOLOGICAL DETERMINANTS OF HIGH BLOOD PRESSURE?

Much of the research linking psychological variables and blood pressure has been done in an effort to discover the "cause" of permanent high blood pressure or hypertension. Such studies were begun on the hunch, and some evidence, that there were behavioral differences between persons with high blood pressure and those with low. For example, Alexander (1902) was among the earliest to note differences between the blood pressures of persons committed to mental hospitals and blood pressures of the general population, with the patients tending to have higher blood pressures. In 1919 Moschowitz drew up a personality description of hypertensives based on observations made on his patients. The hypertensive was described as: overweight, sluggish, slow and ungraceful; tense and irritable; with narrow intellectual horizons and no hobbies; sleeping badly; having lost the ability to

enjoy life, even though financially successful. Moschowitz carefully avoided asserting this syndrome as the cause of high blood pressure. Other examples of this sort of research are studies by Malmö and Shagass (1950) and Benedict (1956).

#### *Specific Personality Differences of High and Low Blood Pressure Groups*

Using a combination of self-ratings and ratings by two peers, Hamilton (1942) found that young high blood pressure subjects are more frequently described as submissive, introverted, less self-assertive, less self-confident, slow, lethargic, and getting less enjoyment from the opposite sex. These results seem similar to those found in a population of college males by Harburg, Julius, McLeod, McGinn, and Hoobler (1963) who report positive correlations of Cattell's 16 PF measures of "sensitivity" and "submissiveness" with consistently high systolic blood pressure levels measured in different situations; for example, on a college registration line, in the physician's office, and at home taken by the student. Higher blood pressure among this population was also associated with yielding in an argument between peers.

Brower (1947a) correlated blood pressure of college students with scores on the Rorschach and MMPI. He found a significant negative correlation between diastolic blood pressure and *FC* scores on the Rorschach—*FC* is considered by Brower to be a measure of adjustment. In another study, Brower (1947b) found significant positive correlations between diastolic blood pressure (he did not measure systolic) and the  $\Sigma$  Depression, Hypochondriasis, and Psychopathic Deviate scales of the MMPI. Cattell and Scheier (1959) report correlations between systolic blood pressure and some of the Cattell 16 PF factors.

No further description is provided of the procedure or the subjects. They claim a positive correlation between high systolic pressure and Extraversion, A+, Q2-, N+, M-, E+, having bigger bones, exuberance, long range goals, low confidence of skill in untried performance, and more tremor in right hand.

There are a few studies which report no differences in scores on personality tests between high and low blood pressure samples. For example, Innes, Millar, and Valentine (1959) found no significant differences between hypertensive pregnant females (average age 26), normotensive females after delivery (age 26), psychoneurotic females (age 32), and hypertensive females (age 53) in scores on the MMPI. Robinson (1959) found no significant relationship between blood pressure and performance on the Maudsley Personality Inventory and the Porteus Mazes. Storment (1951) compared five groups of male hospital patients using the Guilford-Martin STDCR, GAMIN, and OAgCo. The groups of patients were: hypertensives, rheumatic heart disease, coronary occlusion, arteriosclerosis, and non-cardiac infectious disease patients. There were no differences among the five groups on test scores.

### *Hostility and High Blood Pressure*

One of the most energetic "schools" of research on hypertension is that which promotes the repressed hostility hypothesis. Beginning with Alexander's (1939a) theoretical discussion, this hypothesis asserts that hypertension develops because the prehypertensive person continually represses feelings of hostility instead of expressing them. This repressed hostility or tension is somatically expressed in increased blood pressure which, over time, leads to the permanent vascular changes associated with hypertension. Saul (1939) treated

hypertensives in analysis and concluded that the salient similarities among hypertensives are "the prominence in every case of a dominating mother . . . marked inhibition of heterosexuality, repressed hostility and being neither weak and dependent nor aggressively hostile [p. 160]."

A number of clinical observers have worked along the line of Alexander's repressed hostility hypothesis. They usually interpret the material obtained in psychotherapy sessions. It will remain an open question as to how representative of the general population of hypertensives are those in psychotherapy; it is quite likely that they are a small and selected part of the population. Sample studies in this area, yielding results much like Saul's, are those by Wolf et al. (1948), Gressel, Shobe, Saslow, DuBois, and Schroeder (1949), and Saul, Sheppard, Selby, Sachs, and Master (1954).

There are other studies which raise problems for the idea that hypertensives repress their hostility. Neiberg (1957) gave a multiple-choice form of the Rosenzweig Picture-Frustration Study before and after a criticism experience. There were no differences in poststress aggressivity between hypertensives and normal subjects. Schachter (1957) exposed subjects to pain, fear, and anger situations in that order. Judges rated expression of anger and fear by the subjects. High blood pressure subjects expressed more anger and fear; there was a positive correlative correlation between expression of anger and blood pressure. A study by Kaplan, Gottschalk, Magliocco, Rohorst, and Ross (1961) further suggests that hypertensives are more hostile than those with normal blood pressure, but that they *express* their hostility in measurable ways. Shapiro (1960) has reviewed research relevant for the repressed hostility

ity hypothesis and concludes that available data are inconclusive.

### *The Prehypertensive Personality*

In order to successfully argue that a particular personality syndrome precedes the development of hypertension (and might therefore be causative) it is necessary to demonstrate that persons of a given personality type are more likely to develop hypertension than are others. No such research has been done. Indirect evidence is provided in two studies (Ayman, 1933; Kalis, Harris, Sokolow, & Carpenter, 1957) which show that young subjects with high blood pressure are more likely than normals to develop hypertension (e.g., Levy, White, Stroud, & Hillman, 1945) and are similar in personality to older persons with hypertension.

Several other examples of research lend partial support to the idea that emotional responses, conditioned by early learning, may in turn mediate blood pressure reactivity. An experiment by Hokanson (1961b) provided subjects with a stressing social situation. Increase in systolic and diastolic blood pressures under social stress were correlated with scores on several tests of manifest hostility. Subjects high in Test Hostility showed greater rises in systolic and diastolic blood pressure. Hokanson argues that subjects with high Test Hostility scores must feel anxious or guilty about expressing hostility. He relates this to the findings in another study (1961a) in which subjects who released aggression showed a greater decline in blood pressure after the experiment than did subjects who did not aggress. Hokanson contends that "aggression anxiety" acts to inhibit aggression and is related to early emotional conditioning of the child by his parents.

Some evidence is available to support the argument that blood pressure

reactivity may be genetically transmitted. For example, studies by Hines (1940), Doyle and Fraser (1961), and Shapiro (1961) demonstrate that children of hyperresponsive parents show greater increases of blood pressure to stress than do children of normoresponsive parents. However, a respectable alternative hypothesis, which fits the data equally well, is that hypertensive parents create conditions in the home which lead to development of high blood pressure or reactivity in their children. As far as is known, the only research relating a subject's reported behavior of his parents and his own blood pressure reactivity is that by McGinn (1962; McGinn, Harburg, & Julius, in press). He found that hyperresponsive college males described their mothers as aggressively rejecting them, and fathers passively rejecting, as compared with normoresponsive subjects. These data are consistent with the hypothesis that stressful experiences in childhood can lead to physiological overreactivity to stress in adulthood. Two descriptions of the process by which this might occur are those by Mowrer and Viek (1948) and Malmo (1957). However, there are no studies which have employed direct observation of parental behavior rather than children's or parents' recall.

Finally, if personality characteristics precede development of hypertension and are causally associated with it, then one might expect to find different personality traits among essential hypertension patients as compared with those whose hypertension can be ascribed to some structural malfunction, such as a dysfunctional kidney. However, Ostfield and Lebovits (1958) found that essential hypertension patients and renal hypertension patients, matched on age, sex, race, and socioeconomic status, showed no differences in group scores on the MMPI or Rorschach or in response to

a stressful interview. It is always a possibility that personality characteristics are a noncausal correlate of hypertension in all stages of its development and that both can be attributed to some other causal agent.

### *Effects of Environmental Stress*

One such possible agent is environmental stress. Inadequate behavior in social situations might have disturbing consequences for some people, hence the relationship of personality measures with blood pressure. For other persons, occupational strains or other stressing events in the daily environment might occasion frequent elevation of blood pressure and lead to the development of the illness in the way suggested by Malmö (1957).

Steiglitz (1930) was perhaps the first to argue that repeated episodes of emotional tension could lead to permanent hypertension. Palmer (1937) performed a study with the cold pressor test and concluded that hypertension is a chronic emergency response:

Those who have definite essential hypertension usually have experienced severe and prolonged mental or nervous strain and often have suffered from insomnia, functional gastric disorder, marked exhaustion and occasionally, more or less severe nervous breakdown [p. 692].

Reiser and his colleagues (Reiser, Brust, & Ferris, 1951; Reiser, Brust, Shapiro, Baker, Ranschoff, & Ferris, 1950; Reiser, Rosenbaum, & Ferris, 1951) have attempted to link the occurrence of life crises and the development of hypertension by intensive clinical examination of the events which preceded the onset of the illness. No specific stress seemed a necessary condition, but the authors feel these stresses usually represent threats to the patient's self-concept, such as threatened divorce, or failure on the job.

Another way to test the environmental

stress hypothesis is to examine blood pressures across cultures, predicting that cultures characterized by highly stressful conditions of life will have more hypertensives than those cultures in which life is easier and less stressful. For example, Shreenivas (1951) found lower blood pressures among natives of East India than are found in the United States. Maddocks (1961a, 1961b) found lower blood pressures in South Sea Islanders than in residents of London, England. However, why should it be expected that the overcrowded conditions of life in the Orient are less stressful than those in more advanced economies? Further, Marcussen (1950) discusses several cases in which symptoms of high blood pressure disappeared when the subject was under an extreme environmental stress (e.g., being a prisoner of war of the Japanese) but reappeared upon re-establishment of a normal pattern of life.

The most effective test of the stress-produces-hypertension hypothesis would be to expose an organism to stress and see whether it develops hypertension. Selye (1943); Farris, Yeakel, and Medoff (1945); and Yeakel, Shenkin, Rothballer, and McCann (1948) claim that rats exposed to loud noises (sounds produced by an air blast) at frequent intervals for a long period of time develop hypertension. Russian researchers have also been able to experimentally produce hypertension in animals using Pavlovian conditioning techniques, according to a review of their work by Simonson and Brozek (1959).

However, some researchers have not been able to duplicate these results. Shapiro and Horn (1955) used conditioning techniques in an attempt to produce experimental neuroses in cats. Full-blown neurotic behavior never developed, according to the authors. Neither was the blood pressure of the conditioned animals markedly above

that of controls. Shapiro and Melhado (1958) exposed rats to various conditioning procedures designed to evoke "chronic anxiety." No hypertension was produced, but there was evidence that existing vascular disease could be seriously exacerbated by the stress procedures. The Shapiro studies used a conditioning technique rather than the noxious stimuli procedures of Selye and Farris because, as Shapiro (1960) has pointed out, the latter involve long periods of extreme fatigue for the animal in addition to "psychic" effects, thus clouding the interpretation of results. On the other hand, there is reason to suspect that Shapiro's animals never experienced the degree of stress that the Selye and Farris animals did and that this amount of stress is necessary to produce permanent high blood pressure.

### DISCUSSION

In conclusion, it should be noted that most of the extant research is the product of members of the medical profession. For some unclear reason psychologists lost interest in blood pressure as an important physiological variable to be related to personality factors shortly before 1930. Some valuable research has been produced by psychologists in very recent years, but in small quantity. One of the consequences of domination of the field by medical researchers is that much of the research does not meet those methodological criteria which psychologists would employ.

The following problems are seen as those which are most in need of solution at the present time. (a) Can permanent high blood pressure be induced in animals by techniques which are commonly accepted as producing experimental neuroses? (b) What, if any, is the relationship between childhood experiences and blood pressure characteristics? (c) Are certain young adult per-

sonality types more likely to develop high blood pressure than others?

### REFERENCES

- ALEXANDER, F. Emotional factors in essential hypertension: Presentation of a tentative hypothesis. *Psychosom. Med.*, 1939, 1, 175-179. (a)
- ALEXANDER, F. Psychoanalytic study of a case of essential hypertension. *Psychosom. Med.*, 1939, 1, 139-152. (b)
- ALEXANDER, H. A few observations on the blood pressure in mental diseases. *Lancet*, 1902, No. 2, 18-20.
- ANGELL, J. R., & THOMPSON, H. B. Organic processes and consciousness. *Psychol. Rev.*, 1899, 6, 32-69.
- AX, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, 15, 433-442.
- AYMAN, D. Personality type of patients with arteriolar essential hypertension. *Amer. J. med. Sci.*, 1933, 186, 213-223.
- BARD, P. (Ed.) *Medical physiology*. St. Louis, Mo.: Mosby, 1956.
- BENEDICT, RUTH B. Onset and early course of essential hypertension. *J. chron. Dis.*, 1956, 4, 221-230.
- BROWER, D. The relation between certain Rorschach factors and cardiovascular activity before and after visuo-motor conflict. *J. gen. Psychol.*, 1947, 37, 93-95. (a)
- BROWER, D. The relations between Minnesota Multiphasic Personality Inventory scores and cardiovascular measures before and after experimentally induced visuo-motor conflict. *J. soc. Psychol.*, 1947, 26, 55-60. (b)
- BUSS, A. H. *The psychology of aggression*. New York: Wiley, 1961.
- CANNON, W. B. *The wisdom of the body*. New York: Norton, 1932.
- CATTELL, R. B., & SCHEIER, I. H. Extension of meaning of objective test personality factors: Especially into anxiety, neuroticism, questionnaire, and physical factors. *J. gen. Psychol.*, 1959, 61, 287-315.
- CRANSTON, R. W., CHALMERS, J. H., TAYLOR, H. L., HENSCHEL, A., & KEYS, A. Effect of a psychiatric interview on the blood pressure response to cold stimuli. *Federat. Proc.*, 1949, 8, 30.
- DARROW, C. W. Differences in the physiological reactions to sensory and ideational stimuli. *Psychol. Bull.*, 1929, 26, 185-201.
- DOYLE, A. E., & FRASER, J. R. E. Essential hypertension and inheritance of vascular reactivity. *Lancet*, 1961, No. 2, 509-511.
- FARRIS, E. J., YEAKEL, D. H., & MEDOFF, H. W. Development of hypertension in emotional

- gray Norway rats after air blasting. *Amer. J. Physiol.*, 1945, 144, 331-341.
- FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, MARGARET E. *The mastery of stress*. Cambridge: Harvard Univ. Press, 1957.
- GRAHAM, J. D. P. High blood pressure after battle. *Lancet*, 1945, No. 1, 239-240.
- GRESSEL, G. C., SHOBE, F. O., SASLOW, G., DuBOIS, P. J., & SCHROEDER, H. A. Personality factors in essential hypertension. *J. Amer. Med. Ass.*, 1949, 140, 265-271.
- GRIMAK, L. P. Reproduktsiia emotsional'nikh sostoianii parashiotista v. gipnoze. [Reproduction of emotional states of a parachutist under hypnosis.] *Vop. Psikh.*, 1959, 5(3), 139-142.
- HAMILTON, J. A. Psychophysiology of blood pressure: I. Personality and behavior ratings. *Psychosom. Med.*, 1942, 4, 125-133.
- HARBURG, E. Covert hostility: Its social origins and relationship to overt compliance. Unpublished doctoral dissertation, University of Michigan, 1962.
- HARBURG, E., JULIUS, S., McLEOD, J. H., McGINN, N. F., & HOOBLER, S. W. Personality traits and behavioral patterns associated with systolic blood pressure levels in college males. *J. chron. Dis.*, in press.
- HINES, E. A., JR. Range of normal blood pressures and subsequent development of hypertension. *J. Amer. Med. Ass.*, 1940, 115, 271-274.
- HOKANSON, J. E. The effects of frustration and anxiety on overt aggression. *J. abnorm. soc. Psychol.*, 1961, 62, 346-351. (a)
- HOKANSON, J. E. Vascular and psychogalvanic effects of experimentally aroused anger. *J. Pers.*, 1961, 29, 30-39. (b)
- INNES, G., MILLAR, W. M., & VALENTINE, M. Emotion and blood pressure. *J. ment. Sci.*, 1959, 105, 840-851.
- KALIS, B. L., HARRIS, R. E., SOKOLOW, M., & CARPENTER, L. G., JR. Response to psychological stress in patients with essential hypertension. *Amer. Heart J.*, 1957, 53, 572-578.
- KAPLAN, S., GOTTSCHALK, L. A., MAGLIOCCO, E. B., ROHORST, D. D., & ROSS, W. D. Hostility in verbal production and hypnotic dreams of hypertensive patients. *Psychosom. Med.*, 1961, 23, 311-322.
- KOLLER, E. K., & KATZ, J. (Eds.) *A bibliography of the world literature on blood pressure: 1920-1950*. Boston: Commonwealth of Massachusetts, 1952.
- LANGFELD, H. S. Psychological symptoms of deception. *J. abnorm. Psychol.*, 1920-21, 15, 319-328.
- LENNARD, H. L., & GLOCK, C. Y. Studies in hypertension: VI. Differences in the distribution of hypertension in Negroes and whites: An appraisal. *J. chron. Dis.*, 1957, 5, 186-196.
- LEVY, R. L., WHITE, P. D., STROUD, W. D., & HILLMAN, C. D. Transient hypertension: The relative prognostic importance of various systolic and diastolic levels. *J. Amer. Med. Ass.*, 1945, 128, 1059-1061.
- LOVELL, G. D. Physiological and motor responses to a regularly recurring sound. *Psychol. Bull.*, 1941, 38, 715. (Abstract)
- McGINN, N. F. Perception of parents and blood pressure. Unpublished doctoral dissertation, University of Michigan, 1962.
- McGINN, N. F., HARBURG, E., & JULIUS, S. Blood pressure reactivity and recall of treatment by parents. *J. abnorm. soc. Psychol.*, in press.
- MACWILLIAMS, J. A. Some applications of physiology to medicine: III. Blood pressure and heart action in sleep and dreams. *Brit. med. J.*, 1923, 2, 1196-1200.
- MADDOCKS, I. The influence of standard of living on blood pressure in Fiji. *Circulation*, 1961, 24, 1220-1223. (a)
- MADDOCKS, I. Possible absence of essential hypertension in two complete Pacific island populations. *Lancet*, 1961, No. 2, 396-399. (b)
- MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, 64, 276-287.
- MALMO, R. B., & SHAGASS, C. Studies of blood pressure in psychiatric patients under stress. *Psychosom. Med.*, 1952, 14, 82-93.
- MARCUSSEN, R. M. Vascular headache experimentally induced by presentation of pertinent life experience: Modification of the course of vascular headache by alterations of situations and reactions. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1950, 29, 609-614.
- MARSTON, W. M. Systolic blood pressure symptoms of deception. *J. exp. Psychol.*, 1917, 2, 117-163.
- MOSCHOWITZ, E. Hypertension: Its significance, relation to arteriosclerosis and nephritis, and etiology. *Amer. J. med. Sci.*, 1919, 158, 668-684.
- MOSES, L., DANIELS, G. E., & NICKERSON, J. L. Psychogenic factors in essential hypertension: Methodology and preliminary report. *Psychosom. Med.*, 1956, 18, 471-485.
- MOWRER, O. H., & VIEK, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, 43, 193-200.
- NEIBERG, N. A. The effects of induced stress on the management of hostility in essential hypertension. *Dissert. Abstr.*, 1957, 17, 1597.
- O'HARE, J. P. Vascular reactions in vascular hypertension. *Amer. J. med. Sci.*, 1920, 159, 369-380.

- OSTFIELD, A., & LEBOVITS, B. Z. A comparison of the effects of psychological stress in renal and essential hypertension. *Psychosom. Med.*, 1958, 20, 444. (Abstract)
- PALMER, R. S. The factor of mental stress in essential hypertension. *N. Engl. J. Med.*, 1937, 216, 689-693.
- PALMER, R. S. Psyche and blood pressure. *J. Amer. Med. Ass.*, 1950, 144, 295-298.
- PFEIFFER, J. B., JR., & WOLFF, H. G. Studies in renal circulation during periods of life stress and accompanying emotional reactions in subjects with and without essential hypertension: Observations on the role of neural activity in the regulation of renal blood flow. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1950, 29, 929-953.
- REISER, M. F., BRUST, A. A., & FERRIS, E. B. Life situations, emotions, and the course of patients with arterial hypertension. *Psychosom. Med.*, 1951, 13, 133-139.
- REISER, M. F., BRUST, A. A., SHAPIRO, A. P., BAKER, H. M., RANSCHOFF, W., & FERRIS, E. B. Life situations, emotions and the course of patients with arterial hypertension. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1950, 29, 870-880.
- REISER, M. F., ROSENBAUM, M., & FERRIS, E. B. Psychologic mechanisms in malignant hypertension. *Psychosom. Med.*, 1951, 13, 148-159.
- ROBBINS, S. D. A plethysmographic study of shock and stammering. *Amer. J. Physiol.*, 1919, 12, 195-202.
- ROBINSON, J. D. A study of the relationship between blood pressure and certain aspects of personality. *Bull. Brit. Psychol. Soc.*, 1959, 37, 5A. (Abstract)
- RUSKIN, A., BEARD, O. W., SCHAFER, R. L. Blast hypertension: Elevated arterial pressure and victims of the Texas City disaster. *Amer. J. Med.*, 1948, 4, 228-236.
- SAUL, L. A. Hostility in cases of essential hypertension. *Psychosom. Med.*, 1939, 1, 153-161.
- SAUL, L. A., SHEPPARD, EDITH, SELBY, DOROTHY, LHAMON, W., SACHS, D., & MASTER, REGINA. The quantification of hostility in dreams with references to essential hypertension. *Science*, 1954, 119, 382-383.
- SCHACHTER, J. Pain, fear, and anger in hypertensives and normotensives. *Psychosom. Med.*, 1957, 19, 17-29.
- SCHULZE, V. E., & SCHWAB, E. H. Arteriolar hypertension in the American Negro. *Amer. Heart J.*, 1936, 11, 66-74.
- SELYE, H. On the production of malignant hypertension by chronic exposure to various damaging agents. *Rev. Canad. Biol.*, 1943, 2, 501-505.
- SHAPIRO, A. P. Psychophysiological mechanisms in hypertensive vascular disease. *Ann. intern. Med.*, 1960, 53, 64-83.
- SHAPIRO, A. P. An experimental study of comparative responses of blood pressure to different noxious stimuli. *J. chron. Dis.*, 1961, 13, 293-311.
- SHAPIRO, A. P., & HORN, P. W. Blood pressure, plasma pepsinogen, and behavior in cats subjected to experimental production of anxiety. *J. nerv. ment. Dis.*, 1955, 122, 222-231.
- SHAPIRO, A. P., & MELHADO, J. Observations on blood pressure and other physiologic and biochemical mechanisms in rats with behavioral disturbances. *Psychosom. Med.*, 1958, 20, 303-313.
- SHREENIVAS, M. B. B. S. Blood pressure—normal range. *Indian Heart J.*, 1951, 3, 22-26.
- SIMONSON, E., & BROZEK, J. Review: Russian research on arterial hypertension. *Ann. intern. Med.*, 1959, 50, 129-193.
- STEIGLITZ, E. J. Emotional hypertension. *Amer. J. med. Sci.*, 1930, 179, 175-182.
- STEINMAN, B., JAGGI, U., & WIDMER, J. Über den Einfluss von Geräuschen und Lärm auf den Blutdruck des Menschen. *Cardiologia*, 1955, 27, 233-239.
- STEVENSON, I. P., DUNCAN, C. H., FLYNN, J. T., & WOLF, S. Hypertension as a reaction pattern to stress: Correlation of circulatory hemodynamics with changes of attitude and emotional state. *Amer. J. med. Sci.*, 1952, 224, 286-299.
- STORMENT, CHARLYNE TOWNSEND. Personality and heart disease. *Psychosom. Med.*, 1951, 13, 304-313.
- THURRELL, R. J., GREENFIELD, N. W., & ROESSLER, R. The prediction of physiological responsivity from psychological responsivity. *J. psychosom. Res.*, 1961, 5, 211-214.
- WHITE, B. V., JR., & GILDEA, E. F. "Cold pressor test" in tension and anxiety: A cardiocronographic study. *Arch. Neurol. Psychiat.*, 1937, 38, 914-945.
- WOLF, S., PFEIFFER, J. B., RIPLEY, H. S., WINTER, O. S., & WOLFF, H. G. Hypertension as a reaction pattern to stress: Summary of experimental data on variations in blood pressure and renal blood flow. *Ann. intern. Med.*, 1948, 29, 1056-1076.
- YEAKEL, ELEANOR H., SHENKIN, H. A., ROTHBALLER, A. B., & MCCANN, S. M. Blood pressure of rats subjected to auditory stimulation. *Amer. J. Physiol.*, 1948, 155, 118-127.

(Received October 24, 1962)

## ASSESSMENT OF RISK TAKING BEHAVIOR

PAUL SLOVIC<sup>1</sup>

*University of Michigan*

Risk taking propensity has been hypothesized to be a general personality disposition and many devices have been proposed for use in its assessment. Several types of studies relevant to the validity of these measures are reviewed. These studies expose a considerable lack of agreement among what are supposed to be measures of the same general characteristic. This implies that the domain of risk taking behavior may not be as conceptually unitary as many psychologists would like to believe. Some explanations for this lack of agreement are offered in an attempt to provide a more adequate means of conceptualizing and assessing risk taking behavior.

An increasing number of recent theoretical and experimental articles have attempted to provide insights into the dynamics of risk taking (RT) and to relate individual differences in RT to other psychological variables. For example, RT propensities have been hypothesized to be important determiners of problem solving ability (Bruner, Goodnow, & Austin, 1956), creativity (McClelland, 1956a), accidents (Dunlap, 1953; Suchman & Schertzer, 1960), vocational choice and entrepreneurship (McClelland, 1956b, 1961; Ziller, 1957b), and criminality (Cohen & Hansel, 1956). For the most part these studies have simply proposed a relationship between behavior and RT propensity and have been little concerned with the problems inherent in defining and assessing this presumably general personality characteristic. The purpose of this paper is to focus attention on the problem of developing an adequate conception of RT behavior upon which to base assessment devices for application in research of this nature.

### HISTORY AND VALIDITY OF RT MEASURES

Most dictionary definitions of risk center around the phrase "chance of loss." Accordingly, the magnitudes of both probabilities and losses enter into the determination of the amount of risk which a given course of action entails. This section is concerned with acquainting the reader with the rationales by which proposed RT measures are linked with the notions of probability and loss and with integrating three types of evidence relevant to the adequacy with which they are measuring general RT propensities. At this early stage of research, such evidence also has implications for the adequacy of the conception of RT as a behavioral tendency which will consistently exhibit itself in a variety of situations.

The first type of evidence comes from studies attempting to correlate RT indices with other personality traits. If the different assessment devices are tapping a common dimension of personality they should have similar personality correlates. Unfortunately, only two variables, need for achievement and masculinity-femininity, have been studied in enough detail to make this kind of analysis possible. A second type of evidence is obtained from comparison of

<sup>1</sup> Now at Oregon Research Institute, Eugene.

The author wishes to thank Ward Edwards, Sarnoff A. Mednick, H. William Morrison, and Warren T. Norman for their helpful criticisms and suggestions.

the measures with independent "risk relevant" behaviors such as the amount of risk in one's vocation. Finally, valuable information is gained by comparing each instrument with other measures of RT propensities in an attempt to establish what Campbell and Fiske (1959) call convergent validity. Convergent validity requires different measures of the same trait to correlate significantly and substantially with one another. It is a necessary step whenever the trait under study is one which can presumably be consistently observed in different situations.

To facilitate these proposed analyses, RT indices have been divided, somewhat arbitrarily, into three categories: response set and judgmental, questionnaire, and probability and variance preference measures.

#### *Response Set and Judgmental Measures*

A number of investigators have been concerned with willingness to gamble on ability tests as reflective of general RT tendencies. When Cronbach (1946) discussed the types of response sets that contaminate test scores he mentioned several that have been of particular interest to researchers seeking objective measures of RT. These are the tendency to guess on a test item when uncertain about the answer rather than omit it (Cronbach labeled this the gambling responses set); the tendency to include many items in a given category when no limit is imposed (the inclusiveness set); the tendency to answer rapidly on time tests, attempting to compensate for committing a greater number of errors by completing more items (the speed versus accuracy set).

*Gambling Set.* The tendency to gamble by guessing on tests that impose a penalty for mistakes was noted early by Votaw (1936). Swineford (1938, 1941)

had subjects indicate the number of points that they wanted each question on an achievement examination to be worth. Possible values ranged from one to four points and in the event the item was answered incorrectly, the indicated number of points was subtracted from the subject's score. Swineford's gambling index, the number of errors on questions valued at four points divided by the total number of errors and omissions, was found to be independent of ability and higher for boys than for girls. This latter finding is an expected one in our culture and is obtained with several other proposed measures of RT. Coombs, Milholland, and Womer (1956) and Danielson (1957) have also hypothesized a relationship between test gambling and RT, the latter investigator concluding that gambling proneness is a personality variable whose expression can be modified by conditions such as ambiguity or importance of the test. Ziller (1957a, 1957b) constructed an index of test gambling and related it to preferences for risky occupations. Persons aspiring to sales occupations gambled most often while future civil engineers took fewest risks.

*Inclusiveness Set.* The inclusiveness set was first viewed as reflecting a personality trait by Klüver (1936). More recently, Gardner (1953) noted that persons vary greatly in the realm of objects that they are willing to subsume under the same rubric. He found that individual differences in this equivalence range were consistent over a variety of perceptual and nonperceptual tasks. This concurs with findings by Marrs (1955).

Pettigrew (1958) studied a variable called category width that is quite similar to Gardner's notion of equivalence range and to the inclusiveness set. Pettigrew found that when subjects were asked to estimate maximum and mini-

mum values for such items as the length of whales or the annual rainfall in Washington, D. C. they exhibited a significant consistency in their category ranges. Pettigrew suggested that category width may be tapping an RT dimension. For example, broad categorizers seem to have a tolerance for Type I errors: They risk negative instances in an effort to include a maximum of positive instances. By contrast, narrow categorizers are willing to make Type II errors: They exclude many positive instances by restricting their category ranges in order to minimize the number of negative instances.

Bruner and Tajfel (1961) proposed a similar conception of category width in RT terms. However, unless differential costs are explicitly associated with the two types of error, it is not possible to label broad categorizing as necessarily riskier than narrow categorizing or vice versa. Therefore, it is surprising that most research has attempted to link broad categorizing with high RT propensity. Wallach and Kogan (1959) and Wallach and Caron (1959) found that males were broader categorizers than females. Wallach and Caron also found that boys were more willing than girls to call ambiguous figures similar to a standard figure in a situation where the likelihood of error was greater when these figures were judged similar. These similarity judgments were found to be related to category width. Rosen (1961) and Rosen and Hoffman (1961) found that males had greater category widths than females and that subjects reduced their category widths under conditions designed to promote avoidance of risk.

*Speed versus Accuracy.* Persons who sacrifice accuracy for speed and who are willing to gamble on an examination exhibit a readiness to risk making Type I errors similar to that of the broad cate-

gorizer. Johnson (1954) and Block and Petersen (1955) found speed versus accuracy scores to be quite stable across different tasks. Winder and Wurtz (1954) studied personality differences among fast and slow deciders and found that fast deciders scored higher on the Masculinity scale of the MMPI. Fillenbaum (1959) provided evidence indicating moderate generality of category width but failed in an attempt to relate category width variables and decision speed. Only 4 out of 10 correlations among different category width variables were significant, the highest of those being .40. None of the category width measures correlated with decision time as measured by a task in which subjects were asked to estimate the most common characteristic of a deck of cards after observing as few cards as possible.

*Judgmental Measures.* Other potential RT indices include measures of extremity of judgment in situations where greater extremity affords the possibility of a greater magnitude of error and judgmental confidence or certainty, which might be indicative of an individual's characteristic biases in perceiving probabilities of success and failure. Brim and Hoff (1957) investigated the Desire for Certainty Test in which subjects were asked to complete sentences of the form "The chances that such-and-such an event will occur are about \_\_\_\_\_ in 100." After making his probability estimate the subject rated his confidence in that estimate. Scores obtained were the mean confidence rating and mean deviation from the most conservative probability estimate (which is 50%). Wallach and Kogan (1959) studied sex differences on the Desire for Certainty Test. They found that for items on which persons were anything less than quite sure, men's probability estimates were more extreme than those of women. For items on which subjects were very

sure, women's probability estimates were more extreme. Women also had less confidence in their probability estimates and narrower category widths. They concluded that women were highly certain less frequently than men, but when they were very certain they were more willing to take large risks. Similarly, Winder and Wurtz (1954) found that highly confident individuals had higher masculinity scores on the MMPI than did less confident persons and were more likely to describe themselves as adventurous on an adjective check list. Kogan and Wallach (1960) attempted to validate the interpretation of judgment certainty as an RT concept by relating subjects' confidence in their probability estimates on the Desire for Certainty Test to their semantic differential evaluation of concepts assumed to be relevant to general risk, monetary risk, or physical risk. Persons less certain of their judgments tended to conceive of risk taking terms as more hostile, cold, and tense than did more confident individuals.

#### *Questionnaire Measures*

Torrance and Ziller (1957) constructed an inventory to assess RT propensities from a knowledge of life experiences. Their item pool consisted of activities hypothesized to be related to the development of RT tendencies. These items covered early childhood behaviors, sports, reactions to competition, financial risks, risks in interpersonal relationships, etc. They validated their Risk Scale against a questionnaire which required subjects to choose among actions embodying varying degrees of risk in military situations. The indices obtained from this questionnaire were the number of items on which the subject chose the riskiest alternative, the mean probability of success assigned to the riskiest acts, and the minimum probability of success

that the subject said he would require before undertaking the riskiest acts. Correlations between the Risk Scale and these variables ranged from .22 to .34 in the predicted direction. However, when interpreting these results one must note that these questionnaire indices shared a common self-report response modality and variance due to this common method may have inflated the correlations. Correlations between these measures and Ziller's (1957a) index of the gambling set, a measure based on a different response modality, were somewhat lower.

Williams (1960) built a Guttman scale called the Job Preference Inventory with which to measure tendencies to take calculated risks in an industrial setting. People classified as high risk takers were those who stated preferences for jobs in which they: are almost always on their own, have to make many decisions by themselves, could either be highly successful or complete failures, etc. Williams found that workers classified as high risk takers placed more emphasis on promotions and upon merit as the criteria for mobility than did low risk takers. He also discovered that individuals whose jobs provided less challenge in terms of risk than they were willing to accept were more likely to be dissatisfied with their jobs than were workers whose jobs entailed a satisfactory degree of risk.

#### *Probability and Variance Preference Measures*

Edwards (1953, 1954a, 1954b, 1954d) pioneered in the study of preferences among gambles differing in probability of winning and losing. He observed that subjects had definite preferences for betting at some probability levels rather than at others.

Coombs and Pruitt (1960) noted that a bet with probability  $p$  to win  $A$  dollars

and probability  $q$  to win  $B$  dollars can be viewed as a probability distribution over outcomes of money whose mean or expected value is  $p(A) + q(B)$ , whose variance is  $pq(A-B)^2$ , and whose skewness is  $1-2p/\sqrt{pq}$ . For bets with two outcomes, skewness is confounded with the probabilities of winning and losing. Coombs and Pruitt argued that in addition to probability preferences, variance preferences are indicative of an individual's utility for risk. The variance of a bet reflects the amount of deviation of its possible outcomes from the average amount of money to be obtained by playing the bet. For example, a bet offering even odds to win or lose \$5.00 has a larger variance than a bet offering even odds to win or lose \$1.00, although both have an expected value of zero dollars.

Coombs and Pruitt (1960) had subjects choose among pairs of zero expected value bets differing in probability of winning but with variance held constant and vice versa. They found sizable individual differences in probability and variance preferences. Variance preferences were less stable than probability preferences and in general, subjects preferred greater variance for bets which contained their favorite probabilities. The writer (unpublished) replicated the Coombs and Pruitt study and also found subjects who exhibited this shift in variance preference. However, other subjects had variance preferences which were stable at different probability levels and probability preferences which shifted systematically as the variance of the bets changed.

It should be noted that for bets with two outcomes and zero expected value, such as those used by Coombs and Pruitt, alternatives with high probabilities of winning carry with them the possibility of a large loss. Similarly, though long shot bets afford greater opportunities

to lose, the amount of money that may be lost is relatively small. Therefore, there is no unequivocal hypothesis concerning the relation between probability preferences among these bets and high RT. The magnitude of risk which persons perceive in these bets is an open empirical question.

Additional studies, with the exception of Edwards (1954d) and Royden, Suppes, and Walsh (1959), support the notion that variance preferences exist and are consistent for a given individual within a particular RT task. Edwards found that variance preferences were considerably less important than probability preferences in determining choices among bets. Royden et al. attempted to measure subjects' utility for gambling by determining their preferences between a gamble in which they could win either  $a$  or  $c$  cents, the subjective probability of each of these outcomes being .5, and an option which gave them an intermediate value,  $b$  cents, for sure. Subjects who were indifferent between the two options when the  $b$  value was below the expected value of the gamble (in this case the mean of outcomes  $a$  and  $c$ ) were presumed to dislike gambling. The authors reported that a simple dichotomization of persons into those with positive and those with negative utility for gambling was not possible although there was some regularity to their behaviors. In the absence of more detailed presentation of results it is difficult to evaluate this conclusion. Davidson and Marschak (1959), using a procedure quite similar to that of Royden et al., found substantial variance preferences among subjects. In many instances persons chose a bet with relatively low expected value but with variance more suited to their tastes. Lichtenstein (1962), using bets with three outcomes, also found marked variance preferences.

Myers and associates (Myers & Fort, 1961; Myers & Katz, 1962; Myers & Sadler, 1960; Suydam & Myers, 1962) have investigated the effects of variance and other payoff variables on the decision to take risks. In these experiments, subjects were asked to choose between a gamble and a sure thing. The sure thing was sometimes a win and sometimes a loss. They found that as the variance of the gamble increased subjects became more conservative, tending to avoid the larger potential losses more than they approached the equally larger winning outcomes. As expected, subjects gambled more when the alternative to gambling was a certain loss than when it was a sure gain. Unfortunately, no data on individual differences were reported.

Expectancy theorists, especially Atkinson (1957), have stimulated much work related to the study of probability preferences. The impetus for Atkinson's risk taking model came from the relationship that McClelland (1956b, 1958) found between need for achievement and preference for moderate probabilities of success in ring toss, level of aspiration, and vocation choice. The model involves six variables: the subjective probability (i.e., expectancy) of success ( $P_s$ ), the subjective probability of failure ( $P_f$ ), the incentive value of success ( $I_s$ ), the (negative) incentive value of failure ( $I_f$ ), the achievement motive ( $M_s$ ), and the motive to avoid failure ( $M_f$ ). Incentive values and subjective probabilities are assumed to be inversely and linearly related. The variables are combined multiplicatively in the following equation: Resultant Motivation =  $(M_s \times P_s \times I_s) + (M_f \times P_f \times -I_f)$ . The Resultant Motivation function has a maximum at  $P_s = .5$  if  $M_s$  is greater than  $M_f$ , and a minimum at  $P_s = .5$  if  $M_f$  is greater than  $M_s$ . Thus Atkinson predicts that individuals in whom  $M_s$  is greater than  $M_f$  will

prefer tasks with intermediate  $P_s$  while persons dominated by  $M_f$  will prefer tasks in which  $P_s$  is extremely high or low.

Atkinson, Bastian, Earl, and Litwin (1960) found support for the model from the fact that subjects with high  $M_s$  preferred to shoot from moderately difficult distances in a shuffleboard game while high  $M_f$  persons preferred the extremely easy or difficult distances. They also obtained probability preferences among bets in a make-believe gambling situation. Subjects were asked to make choices such as "Would you prefer odds of 1/6 to win \$1,800 or 4/6 to win \$450?" Expected value was held constant but variance decreased monotonically as probability of winning increased. Subjects with low  $M_s$  preferred extreme probabilities but the preferences of subjects with high  $M_s$  were evenly distributed across all probabilities. There was a small positive relationship between a subject's probability preferences in the skill and chance situations. Meyer, Walker, and Litwin (1961) found no relationship between achievement motivation and probability preferences determined from the previously described make-believe bets used by Atkinson et al. Littig (1959) studied preferences among bets in which variance was held constant and obtained results similar to those of Atkinson et al. Persons with high  $M_f$  preferred extreme probabilities but  $S_s$  with high  $M_s$  did not prefer moderate probabilities to extreme ones. Littig attempts to explain this latter finding by suggesting that a chance situation does not arouse achievement aspirations in subjects with high  $M_s$ . Scodel, Minas, and Ratoosh (1959) related probability preferences in a real gambling situation to achievement motivation and other selected personality variables. Variance was not controlled. Subjects preferring intermediate probabilities of success and

intermediate payoffs scored highest on  $M_s$ , supporting Atkinson's theory. In short, studies of RT in achievement-oriented skill situations have generally supported the Atkinson model while probability preferences in strictly chance settings have led to inconsistent results. Several possible explanations for this inconsistency will be offered later.

Both McClelland and Atkinson used level of aspiration tasks as indicators of RT propensities. This line of thought stems from hypotheses presented by Frank (1935) and by Lewin, Dembo, Festinger, and Sears (1944), who believed fear of failure to be an important determiner of such behavior. Moss (1961) defined cautiousness as a general tendency to avoid potential failure or disapproval experiences even though this goal might be attained at the expense of other satisfactions. He found that persons who were cautious on a level of aspiration task also responded cautiously to projective personality measures. That is, they produced descriptive responses to TAT cards and an object sorting task. Such responses afford minimal risk of invoking criticism or personal exposure in contrast to highly idiosyncratic or imaginative protocols.

Another type of measure, closely related to probability preference and cautiousness notions, attempts to determine the amount of information which an individual characteristically requires before committing himself to a major decision based on his knowledge. The probability of making a correct decision increases monotonically with amount of relevant information acquired. Studies by Irwin and Smith (1957), Messick and Hills (1960), Pruitt (1961), and Edwards and Slovic (1962) found stable individual differences in information seeking behavior within a specific type of task. Some subjects consistently bought too much or too little informa-

tion relative to the optimal amount determined by the cost of information and the reward for correct decisions. However, Wolff (1955) failed to find any consistency in behavior across several different information seeking tasks in which costs and payoffs were not explicitly controlled.

Wallach and Kogan (1959, 1961) developed a questionnaire to obtain probability preferences in everyday life situations. Labeled the Dilemmas of Choice Questionnaire in one study and Disutility of Failure Index in another, this measure was studied in terms of sex and age differences. On this test a subject is presented with 12 hypothetical situations, each requiring a choice between a safe alternative and a more attractive but risky one. His task is to indicate the probability of success which would be sufficient for him to select the risky alternative. Men were found to require lower probabilities of success than women for items involving risks of death, income loss, and a football defeat, while women were less conservative than men when the risks involved were those of an uncertain artistic career and a potentially unhappy marriage. Older persons of both sexes required higher probability of success before saying they would undertake the risky act. The internal consistency of required probability of success across the 12 situations was higher for older persons, suggesting that preferred probabilities of success become more generalized with age.

Wallach and Kogan (1961) found that the Disutility of Failure Index was significantly related to a person's subjective probability of failure in a task requiring him to guide a toy car between two posts set at varying distances apart. They concluded that the presence of such a relationship provides a warrant for thinking in terms of greater or less caution as a dimension of personality.

They also attempted, without success, to relate the Disutility of Failure Index to measures of judgmental RT such as confidence and extremity scores on an adaptation of Brim's Desire for Certainty Test. This finding led them to state, contrary to their conclusion in the Kogan and Wallach (1960) study, that until further careful empirical work is carried out in these areas, attempts to conceptualize various judgmental and cognitive processes in decision and risk terms must be considered premature.

An extensive validation study encompassing several different categories of risk taking measures was undertaken by Slovic (1962). Slovic gave subjects tests measuring category width, speed versus accuracy and gambling response sets, William's Job Preference Inventory, Torrance and Ziller's Risk Scale, a vocabulary test in which subjects adjusted the number of points they were willing to risk winning or losing on each item thus controlling the variance of the situation, and a test designed to assess probability and variance preferences.

In addition, subjects were asked to rate their fellow fraternity brothers on a bipolar trait of general willingness to take risks. One pole was defined "Loves to take risks. A dare-devil." The other pole was labeled "Cautious. Does not like to take chances. Avoids risky situations." Twenty-eight correlations were obtained among measures presumed to be positively related. Of these correlations, only five were significant, forcing the conclusion that either very few of the variables measure general RT propensities or such propensities do not exist.

#### DISCUSSION

It is not easy to summarize the many studies concerned with the validity of instruments designed to assess RT. There are enough positive results to

encourage the investigator to further pursue the notion of RT propensity as a general disposition and one which may provide insights into a variety of behaviors. For example, sex differences or scores on tests of masculinity-femininity generally correlate in the expected direction with the different indices of RT. Also, every type of measure is favorably related to some sort of independent behavioral criterion of RT in at least one study. Finally, behavior has been shown to be fairly consistent across measures which involve similar modes of response and are related to RT by a similar rationale.

However, a large amount of evidence bearing on the convergent validity of these methods is negative. The lack of communality across measures involving different response modalities and rationales suggests that future research on promising hypotheses, if undertaken with this heterogeneous collection of assessment devices, will most likely produce inconsistent results. What is needed, therefore, is a systematic investigation of the factors responsible for this lack of convergent validity. Some ideas bearing on this matter will now be discussed.

*Multidimensionality of Risk.* One explanation for the lack of consistency among many RT measures is that risk is a multidimensional concept and most of the presumably "risk relevant" measures have been tapping these dimensions differentially. As mentioned earlier, the usual dictionary definitions of risk center around the expression "chance of loss." Therefore, both probabilities and magnitudes of loss enter into the determination of the amount of risk in a situation. Coombs and Pruitt (1960) took both of these factors into consideration when they conceptualized risk in terms of skewness (probability) and variance. They studied choices among bets in

which these two components were isolated and controlled and found, as did the writer (unpublished), a marked interaction between preferences and the position of stimuli on these dimensions. For example, a person's variance preferences often shifted drastically when the probability level of the bets was changed. This implies that it may be misleading to use variance preferences as an index of RT propensities unless one also considers the probabilities of the stimuli involved. Furthermore, the Coombs and Pruitt study used only bets that had zero expected value. One might argue that a third component of the probability distribution which might characterize the degree of risk in a situation is expected value, since it describes the mean desirability of the outcomes. No one has fully explored preferences among gambles in which expected value, variance, and probability have all been systematically manipulated, but it would seem likely that a complete description of person's RT propensities would require consideration of his unique pattern of preferences in such situations.

*Subjectivity of Risk.* Observation of choices made among acts entailing various degrees of objective risk cannot distinguish the extent to which RT is determined by individual differences in perception of risk or by differences in reaction to that perceived risk. Perceptual differences might take the form of discrepancies among persons' subjective probabilities and subjective values (utilities) and differences in the manner that perceived values on these component dimensions are weighted and combined to produce a resultant evaluation of risk. As if things are not already complicated enough, other parameters of the probability distribution generated by a bet or a real life course of action most likely serve as important cues for the perception of risk. One of these is the

magnitude of the largest possible loss. Another is the proportion of potential outcomes that have negative utility.

The importance of subjective factors in RT is illustrated by familiar circus acts in which tightrope walkers and lion tamers perform their duties with a non-chalance arising from a realistic knowledge of the probabilities of injury, amidst excited crowds who, projecting their own subjective probabilities of failure, feel that the performers are taking great risks. The subjective aspect of RT might well explain the lack of "risk relevance" shown by peer ratings and by response set and judgmental measures for tasks in which subjective probabilities and utilities vary widely between individuals. Researchers may thus be forced to scale subjective probabilities and utilities in order to allow any orderliness inherent in RT behavior maximum opportunity to emerge. For a review of problems and methods of assessing these variables see Edwards (1954c, 1961a, 1962).

*Emotional Arousal and Risk.* Some interesting insights into the theoretical nature of RT and its assessment problems may be obtained by examining the relationship between RT and autonomic stimulation such as that aroused by the emotions of fear and hope. One currently popular school of thought proposes that the individual is constantly seeking some optimal level of internal excitement. Risk is courted in order to raise the amount of excitation when it drops below the optimal level and avoided when the excitation level becomes excessive (Berlyne, 1960; Leuba, 1955).

Cohen and Hansel (1956) provocatively suggested that risky behaviors are usually entered upon, not with conscious or calculated notions of probabilities, but rather with a sense of danger which may be interpreted as fear. Dunlap et al. (1953), studying risk taking among au-

tomobile drivers, hypothesized that the probability that a driver will undertake a risky act is inversely related to his estimate of the risk involved. Furthermore, they proposed, this estimate of risk should be directly proportional to the driver's level of fear. These notions are compatible with those of Mowrer (1960) whose revised learning theory sees all behavior as being mediated by conditioned emotions, especially the emotions of hope and fear. Specifically, Mowrer argues that hopes and fears are the dynamic conditioners of choice and imply not only probabilities but also the nature and magnitudes of risks.

These hypotheses suggest that the study of RT might be fruitfully pursued by investigating each individual's characteristic autonomic responsiveness in order to determine the manner in which subjective probabilities of success and failure interact with outcome utility to produce the sensation of fear. A next step would be to analyze the relationship between this fear and perceived risk and decisions. Mednick (1958) discusses several individual difference factors which influence the fear response and may thus be relevant to RT. These are differences in characteristic anxiety level, reactivity to threat, and rate of recovery from fear arousal.

Most important for the present discussion are the implications of these emotional arousal hypotheses for the evaluation of RT measures. Specifically, it seems likely that emotional arousal is a prerequisite for excitation of RT propensities and that many negative findings may be due to failure on the part of assessment devices to provide the subjects with a true atmosphere of risk.

Risk taking situations may be dichotomized into those whose outcomes are contingent upon skill, such as ring toss, shuffleboard, etc., and those which are controlled only by chance factors

as in the majority of bet preference experiments. Skill tasks, structured to be achievement oriented, are quite likely to promote ego involvement and emotional arousal. Arousal can be promoted in chance situations either by the perception of the situation as one which requires skill or by the introduction of actual monetary rewards and losses as payoffs contingent upon the consequences of the subjects' decisions.

Several personality variables may influence the degree to which persons will become ego involved in a make-believe, chance oriented situation. It would be quite surprising not to find the need for achievement-fear of failure syndrome among them. Littig (1959) hypothesized that persons with high  $M_f$  are more likely to become ego involved in a game of chance than high  $M_s$  individuals. This, if true, might explain why predictions of probability preferences based on Atkinson's model have been most accurate for high  $M_f$  persons in this type of experimental setting. Another variable which might influence ego involvement in chance situations is one's attitude toward fate (Liverant & Scodel, 1960). Persons characterized by a belief in their ability to order their own destiny should be expected to become more involved in a task where success or failure is contingent only upon chance than individuals who believe their fate is not under their own control.

Several recent studies illustrate the importance of arousing RT tendencies by forcing the individual to commit himself to a choice among alternatives whose positive and negative incentives have real meaning for him. Feather (1959) found that subjects in chance oriented situations expressed preferences for less probable goals but actually chose more probable goals when the choice involved commitment to action. Slovic, Lichtenstein, and Edwards (in press)

compared choices among bets under make-believe and real gambling conditions and found that subjects became bored in the make-believe sessions and attempted to ease the burden of making tedious and difficult choices by adopting strategies which enabled them to choose among bets quickly and with a minimum of thought and effort. Shulman (1961) constructed a questionnaire for the purpose of assessing willingness to take risks in many areas of everyday life. He attempted to validate this test against Swineford's (1938) previously described index of gambling on achievement examinations. Two distinct populations of subjects were used. The first consisted of high school students who had been pupils of the experimenter and were well acquainted with him. The second sample consisted of junior college students who had no previous acquaintance with the experimenter. The correlation between scores on the two independent measures of RT was .81 for the high school sample and .16 for the junior college group. Shulman attempted to explain this discrepancy by hypothesizing that an atmosphere of risk in the achievement test situation was aroused only for the high school students whose reputation with the experimenter was at stake. Finally, Rawson (1961) has shown that self-report statements concerning the morality of activities judged to be wrong but risked for exploitative gain were significantly correlated with actual RT behavior in exploitative situations only when risk was aroused.

It seems evident that lack of adequate risk arousal results in behavior which is unrepresentative of the responses which the individual would make under more motivating conditions. Such distortion is likely to be responsible for much of the lack of convergence found among RT measures.

Eliminating motivational biases by introducing real rewards and losses as consequences of decisions creates another possible source of variance for RT behavior, namely, that due to sequential reinforcement effects. Optimal stimulation theorists such as Berlyne (1960) predict that as consequences of previous decisions modify the relationship between one's present arousal level and his optimal state, RT propensities will fluctuate accordingly, thus producing phenotypically inconsistent behavior. The precise nature of effects of reinforcement upon RT is complex and needs considerable study. Effects of sequence on subjective probabilities are already quite well known under the heading of "gamblers' fallacy" or "negative recency effect" (Cohen & Hansel, 1956; Edwards, 1961b). Myers and Fort (1961) provided evidence that the decision to gamble at any particular moment is influenced by the pattern of previous reinforcements, but surprisingly, Slovic, Lichtenstein, and Edwards (in press) and Lichtenstein (1962) found betting behavior to be quite unaffected by past payoff history.

Most of the literature discussed in this paper ignores sequential effects. This literature belongs to the realm of what Edwards (1961a) has labeled "static" decision theory. It assumes that the behavior of interest is sufficiently stable to be of the same relative strength throughout the test. Since any choice must be imbedded in a sequence of choices, a static model can be only a first approximation at best. Nevertheless, static models are much simpler to apply than their dynamic counterparts and they have been shown to work to some extent. Therefore, present investigators of RT behavior should exploit the potentialities of these models while at the same time keeping a watchful eye upon relevant developments in the fields of

motivation, learning, and dynamic decision processes.

*Conclusion.* Risk taking behavior appears to be multidimensional in nature. It has substantial subjective components and is susceptible to a variety of motivational and other influences. Research attempting to establish the validity and consistency of RT measures has generally neglected these factors. It is, therefore, not surprising that so many contradictory results have been found. It seems quite probable that investigations taking these variables into consideration, both in the analysis of the risks involved in criterion situations and in the design of assessment devices, will discover enough orderliness in RT behavior to successfully undertake the various types of research mentioned in the introduction to this paper.

#### REFERENCES

- ATKINSON, J. W. Motivational determinants of risk taking behavior. *Psychol. Rev.*, 1957, 64, 359-372.
- ATKINSON, J. W., BASTIAN, J. R., EARL, R. W., & LITWIN, G. H. The achievement motive, goal setting, and probability preferences. *J. abnorm. soc. Psychol.*, 1960, 60, 27-36.
- BERLYNE, D. E. *Conflict, arousal, and curiosity*. New York: McGraw-Hill, 1960.
- BLOCK, J., & PETERSEN, P. Some personality correlates of confidence, caution, and speed in a decision situation. *J. abnorm. soc. Psychol.*, 1955, 51, 34-41.
- BRIM, O. G., JR., & HOFF, D. B. Individual and situational differences in desire for certainty. *J. abnorm. soc. Psychol.* 1957, 54, 225-229.
- BRUNER, J. S., GOODNOW, JACQUELINE J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
- BRUNER, J. S., & TAJFEL, H. Cognitive risk and environmental change. *J. abnorm. soc. Psychol.*, 1961, 62, 231-241.
- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, 56, 81-105.
- COHEN, J., & HANSEL, M. *Risk and gambling*. New York: Philosophical Library, 1956.
- COOMBS, C. H., MILHOLLAND, J. E., & WOMER, F. B. The assessment of partial knowledge. *Educ. psychol. Measmt.*, 1956, 16, 13-37.
- COOMBS, C. H., & PRUITT, D. G. Components of risk in decision making: Probability and variance preferences. *J. exp. Psychol.*, 1960, 60, 265-277.
- CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 60, 265-277.
- DANIELSON, L. E. Gambling proneness: Its measurement and expression in examination situations. Unpublished doctoral dissertation, University of Michigan, 1957.
- DAVIDSON, D., & MARSCHAK, J. Experimental tests of a stochastic decision theory. In C. W. Churchman (Ed.), *Measurement: Definitions and theories*. New York: Wiley, 1959.
- DUNLAP, J. W., ORLANSKY, J., JACOBS, H. H., YARNOLD, K. W., KISHLER, J. P., & SCHREIBER, R. J. An analysis of risk and exposure in automobile accidents. Report, 1953, Dunlap and Associates, Stamford, Connecticut, Contract DA 49-007-MD-239, Office of the Surgeon General, Department of the Army.
- EDWARDS, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, 66, 349-364.
- EDWARDS, W. Probability preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, 67, 56-67. (a)
- EDWARDS, W. The reliability of probability preferences. *Amer. J. Psychol.*, 1954, 67, 68-95. (b)
- EDWARDS, W. The theory of decision making. *Psychol. Bull.*, 1954, 51, 380-417. (c)
- EDWARDS, W. Variance preferences in gambling. *Amer. J. Psychol.*, 1954, 67, 441-452. (d)
- EDWARDS, W. Behavioral decision theory. *Annu. Rev. Psychol.*, 1961, 12, 473-498. (a)
- EDWARDS, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, 62, 385-394. (b)
- EDWARDS, W. Subjective probabilities inferred from decisions. *Psychol. Rev.*, 1962, 69, 109-135.
- EDWARDS, W., & SLOVIC, P. Seeking information to reduce the risk of decisions. Paper read at Midwestern Psychological Association, Chicago, May 1962.
- FEATHER, N. Success probability and choice behavior. *J. exp. Psychol.*, 1959, 58, 257-266.
- FILLENBAUM, S. Some stylistic aspects of categorizing behavior. *J. Pers.*, 1959, 27, 187-195.
- FRANK, J. D. Some psychological determinants of level of aspiration behavior. *Amer. J. Psychol.*, 1935, 47, 285-293.

- GARDNER, R. W. Cognitive styles in categorizing behavior. *J. Pers.*, 1953, 22, 214-233.
- IRWIN, F. W., & SMITH, W. A. S. Value, cost, and information as determiners of decision. *J. exp. Psychol.*, 1957, 54, 229-232.
- JOHNSON, L. C. Speed and confidence of judgement as psychological variables. Report, 1954, Stanford University, Department of Psychology, Contract Nonr 225-01, Office of Naval Research.
- KLÜVER, H. The study of personality and the method of equivalent and non-equivalent stimuli. *Charact. Pers.*, 1936, 5, 91-112.
- KOGAN, N., & WALLACH, M. Certainty of judgement and the evaluation of risk. *Psychol. Rep.*, 1960, 6, 207-213.
- LEUBA, C. Toward some integration of learning theories: The concept of optimal stimulation. *Psychol. Rep.*, 1955, 1, 27-33.
- LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944.
- LICHTENSTEIN, SARAH C. Bases for preferences among three outcome bets. Unpublished doctoral dissertation, University of Michigan, 1962.
- LITIG, L. W. The effect of motivation on probability preferences and subjective probability. Unpublished doctoral dissertation, University of Michigan, 1959.
- LIVERANT, S., & SCODEL, A. Internal and external control as determinants of decision making under conditions of risk. *Psychol. Rep.*, 1960, 7, 59-67.
- MCCLELLAND, D. C. The calculated risk: An aspect of creative scientific performance. *Collected Pap. Conf. Scient. Creativ., Salt Lake City*, 1956. (a)
- MCCLELLAND, D. C. Interest in risky occupations among subjects with high achievement motivation. Unpublished manuscript, Harvard University, 1956. (b)
- MCCLELLAND, D. C. Risk taking in children with high and low need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.
- MCCLELLAND, D. C. *The achieving society*. Princeton: Van Nostrand, 1961.
- MARRS, C. L. Categorizing behavior elicited by a variety of stimuli. Unpublished master's thesis, University of Kansas, 1955.
- MEDNICK, S. A. A learning theory approach to schizophrenia. *Psychol. Bull.*, 1958, 55, 316-325.
- MESSICK, S., & HILLS, J. Objective measurement of personality: Cautiousness and intolerance of ambiguity. *Educ. psychol. Measmt.*, 1960, 20, 685-698.
- MEYER, H. H., WALKER, W. B., & LITWIN, G. H. Motive patterns and risk performance associated with entrepreneurship. *J. abnorm. soc. Psychol.*, 1961, 63, 570-574.
- MOSS, H. A. The influences of personality and situationally induced cautiousness on conceptual behavior. *J. abnorm. soc. Psychol.*, 1961, 63, 629-635.
- MOWRER, O. H. *Learning theory and the symbolic processes*. New York: Wiley, 1960.
- MYERS, J. L., & FORT, JANE C. A sequential analysis of gambling behavior. Paper read at Psychonomic Society Meetings, New York, September 1961.
- MYERS, J. L., & KATZ, L. Range of payoffs and feedback in risk taking. *Psychol. Rep.*, 1962, 10, 483-486.
- MYERS, J. L., & SADLER, E. Effects of range of payoffs as a variable in risk-taking. *J. exp. Psychol.*, 1960, 60, 306-309.
- PETTIGREW, T. F. The measurement and correlates of category width as a cognitive variable. *J. Pers.*, 1958, 26, 532-544.
- PRUITT, D. G. Informational requirements in risky decisions. *Amer. J. Psychol.*, 1961, 74, 433-439.
- RAWSON, H. E. The relationship of moral value dimensions and unethical behavior under varying conditions of risk. Unpublished doctoral dissertation, Ohio State University, 1961.
- ROSEN, S. Postdecision affinity for incompatible information. *J. abnorm. soc. Psychol.*, 1961, 63, 188-190.
- ROSEN, S., & HOFFMAN, MARGARET F. Effects of risk avoidance on category width and decision certainty. Paper read at Midwest Psychological Association, Chicago, May 1961.
- ROYDEN, H. L., SUPPES, P., & WALSH, K. A model for the experimental measurement of the utility of gambling. *Behav. Sci.*, 1959, 4, 11-18.
- SCODEL, A., MINAS, J., & RATOOSH, P. Some personality correlates of decision making under conditions of risk. *Behav. Sci.*, 1959, 4, 19-28.
- SHULMAN, L. Development of an instrument for the measurement of risk taking tendencies. Unpublished manuscript, University of Chicago, 1961.
- SLOVIC, P. Convergent validation of risk taking measures. *J. abnorm. soc. Psychol.*, 1962, 65, 68-71.
- SLOVIC, P., LICHTENSTEIN, SARAH C., & EDWARDS, W. Boredom induced changes in

- preferences among bets. *Amer. J. Psychol.*, in press.
- SUCHMAN, E. A., & SCHERTZER, A. L. *Current research in childhood accidents*. New York: Association for the Aid of Crippled Children, 1960.
- SUYDAM, MARY M., & MYERS, J. L. Some parameters of risk taking behavior. *Psychol. Rep.*, 1962, 10, 559-562.
- SWINEFORD, F. The measurement of a personality trait. *J. educ. Psychol.*, 1938, 29, 295-300.
- SWINEFORD, F. Analysis of personality trait. *J. educ. Psychol.*, 1941, 32, 438-444.
- TORRANCE, E. P., & ZILLER, R. C. Risk and life experience: Development of a scale for measuring risk taking tendencies. *USAF PTRC res. Rep.*, 1957, No. 57-23.
- VOTAW, O. F. The effect of do not guess directions on the validity of true-false or multiple-choice tests. *J. educ. Psychol.*, 1936, 27, 698-703.
- WALLACH, M. A., & CARON, A. J. Attribute criterionity and sex linked conservatism as determinants of psychological similarity. *J. abnorm. soc. Psychol.*, 1959, 59, 43-50.
- WALLACH, M. A., & KOGAN, N. Sex differences and judgement processes. *J. Pers.*, 1959, 27, 555-564.
- WALLACH, M. A., & KOGAN, N. Aspects of judgement and decision making: Interrelationships and changes with age. *Behav. Sci.*, 1961, 6, 23-36.
- WILLIAMS, L. K. The measurement of risk taking propensity in an industrial setting. Unpublished doctoral dissertation, University of Michigan, 1960.
- WINDER, C. L., & WURTZ, K. R. A study of personality correlates of judgement behavior. Report, 1954, Stanford University, Department of Psychiatry, Contract Nonr 225-01, Office of Naval Research.
- WOLFF, W. M. Certainty: Generality and relation to manifest anxiety. *J. abnorm. soc. Psychol.*, 1955, 50, 59-84.
- ZILLER, R. C. A measure of the gambling response set in objective test. *Psychometrika*, 1957, 22, 289-292. (a)
- ZILLER, R. C. Vocational choice and utility for risk. *J. counsel. Psychol.*, 1957, 4, 61-64. (b)

(Received October 31, 1962)

## NOTE ON FURCHTGOTT'S "BEHAVIORAL EFFECTS OF IONIZING RADIATIONS: 1955-61"

JOHN GARCIA

AND

NATHANIEL A. BUCHWALD

*Long Beach State College*

*University of California, Los Angeles*

Extremely low doses of ionizing radiation operate as complex stimuli with both cue and aversive properties. Recent studies have demonstrated that effects can be obtained with total doses of less than 1 roentgen. This fact may explain some of the controversies concerning radiosensitivity of mammals and the conflicting results obtained in behavioral and electroencephalographic tests for effects of radiation exposure recently reviewed in this journal.

Furchtgott (1963) has performed an excellent service with his recent comprehensive review of the radiation literature. However, the implications of the fact that ionizing radiation is a stimulus were not amply discussed. This fact poses some difficult problems for investigators of behavioral effects and sheds light upon some of the controversial reports emanating from Russian laboratories concerning the effects of extremely low doses. A total dose of 10 roentgens delivered over 4 hours has been employed as an unconditioned stimulus to establish avoidance responses in rats (Garcia, Kimeldorf, & Hunt, 1961). In the first place, this means that the adult animal is more radiosensitive than had been previously believed on the basis of physiological as well as histological examinations. Whether or not these effects are reversible is another question. Furthermore, it means that any stimulus associated with radiation exposure, including handling by the experimenter, may acquire aversive properties and, hence, through generalization, may influence subsequent comparisons of irradiated and nonirradiated animals (Garcia, Kimeldorf, & Hunt, 1956). In some cases postirradiation effects, which are attributed directly to the radiation insult, may be actually conditioned responses to such stimuli by an animal which has completely re-

covered. Conditioned aversions may endure many weeks after a single exposure (Garcia, Kimeldorf, & Koelling, 1955). The contradictory results, increased variability and fluctuating changes in performance, which characterize behavioral data, may be in part due to failure to control this factor. In the same way, physiological measures on unanesthetized intact animals can also be complicated, especially those measures which reflect the general adaptation syndrome described by Selye (1956).

There is no reason to suppose that radiation is an "unconscious" stimulus or "a UCS of which the animal is not aware." Such terms are not only confusing but also misleading. The immediate arousal properties of radiation have been demonstrated in invertebrates; furthermore, immediate visual effects in the dark-adapted eye are well established, as Furchtgott points out. Hence it should not be surprising to find immediate effects upon the electroencephalogram (EEG) or an inhibitory effect upon conditioned responses under certain conditions. Some of the "depressions" of cortical EEG activity reported by Russian investigators resemble the pattern of low voltage, fast activity which accompanies arousal (Livanov, 1962; Tsypin & Grigor'ev, 1960).

In our own laboratory, we have observed the arousal properties of X ray in mammals and also have demonstrated its use as a conditioned stimulus (CS). Unfortunately these most recent data were not yet available to Furchtgott. In sleepy animals X rays produce immediate EEG "desynchronization," that is, the change from high voltage slow waves to low voltage fast waves, as well as behavioral arousal (Garcia, Buchwald, Bach-y-Rita, Feder, & Koelling, 1963). Furthermore, X rays produce an orienting response in active animals which will momentarily interrupt bar pressing in a Skinner box. X rays have been employed as a warning stimulus (CS) to signal a subsequent shock to the animal's paws. After a few trials the animals attempt to escape from the compartment at the onset of radiation indicating that the defensive responses to shock have been conditioned to X ray. In subsequent tests, X rays were employed to inhibit bar pressing with a variation of the Skinner-Estes conditioned suppression technique (Garcia, Buchwald, Feder, & Koelling, 1962).

Radiation is a complex stimulus producing its effects through many different mechanisms. Blind animals respond to X rays, indicating that the retina which is extremely radiosensitive is not necessary for arousal nor avoidance conditioning. Furthermore, the arousal and cue properties differ markedly from the aversive properties in latency and threshold, indicating that different mechanisms may be responsible. Avoidance has been established with intensities as low as .0007 roentgen per second when the exposure was maintained for hours. While .1-.2 roentgen per second will cause arousal within 1 second, the animals do not attempt to escape immedi-

ately; on the contrary they apparently habituate readily to brief exposures.

These apparent discrepancies are probably due to the widespread systemic effects of ionizing radiation which operate through diverse humoral and neural factors to produce their effects upon behavior. Accumulations of breakdown products could account for the aversion effects which appear to be a function of total dose, while the arousal and cue effects seem to indicate a fast responding receptorlike action functionally related to intensity or dose rate.

#### REFERENCES

- FURCHTGOTT, E. Behavioral effects of ionizing radiations: 1955-61. *Psychol. Bull.*, 1963, 60, 157-199.
- GARCIA, J., BUCHWALD, N. A., BACH-Y-RITA, G., FEDER, B. H., & KOELLING, R. A. Electroencephalographic responses to ionizing radiation. *Science*, 1963, 140, 289-290.
- GARCIA, J., BUCHWALD, N. A., FEDER, B. H., & KOELLING, R. A. Immediate detection of x-rays by the rat. *Nature, London*, 1962, 196, 1014-1015.
- GARCIA, J., KIMELDORF, D. J., & HUNT, E. L. Conditioned responses to manipulative procedures resulting from exposure to gamma radiation. *Radiat. Res.*, 1956, 5, 79-87.
- GARCIA, J., KIMELDORF, D. J., & HUNT, E. L. The use of ionizing radiation as a motivating stimulus. *Psychol. Rev.*, 1961, 68, 383-395.
- GARCIA, J., KIMELDORF, D. J., & KOELLING, R. A. Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science*, 1955, 122, 157-158.
- LIVANOV, M. N. The sensitivity of the nervous system of low-level radiation. In, *Effects of ionizing radiation on the nervous system*. Vienna, Austria: The International Atomic Energy Agency, 1962.
- SELYE, H. *The stress of life*. New York: McGraw-Hill, 1956.
- TSYPIN, A. B., & GRIGOR'EV, YU. G. Quantitative measurements of the sensitivity of the central nervous system to ionizing radiation. *Bull. exp. Biol. Med.*, 1960, 49, 26-28.

(Received April 29, 1963)

## POSTSCRIPT TO "BEHAVIORAL EFFECTS OF IONIZING RADIATIONS: 1955-61"<sup>1</sup>

ERNEST FURCHTGOTT

*University of Tennessee*

A reply to Garcia and Buchwald's elaboration on the stimulus characteristics of radiation. (a) Whether radiation should be considered as an unconscious stimulus is a semantic problem. (b) The arousal properties of radiation which have been previously postulated were, however, experimentally demonstrated during the period that the review paper was in press. (c) To account for the diverse Soviet conditioning data in terms of the short-term arousal phenomenon is much too simplified an explanation.

Garcia and Buchwald's (1964) comments elaborating on the treatment of irradiation as a stimulus deal essentially with the following problems:

First, they object to calling radiation an "unconscious" stimulus. In reviewing Arbit's (1958) theoretical paper I simply paraphrased his description of radiation as a stimulus of which the animal may not be "aware." Razran (1961) in his comprehensive review of Soviet studies on interoceptive conditioning says:

Unlike the continuum of exteroceptive stimulation which is the body-material of all our conscious experience, the continuum for interoceptive stimulation leads largely to unconscious reaction [p. 97].

Several studies, including those of Garcia and his co-workers (see Garcia, Kimeldorf, & Hunt, 1961, for a review), as well as a recent study by Barnes (1962) have shown that, in conditioning, radiation acts primarily through the interoceptors and that exclusion of visual and olfactory receptors does not affect the phenomena in which radiation acts as the unconditioned stimulus.

Whether one wants to refer to radiation as unconscious stimulus depends upon one's definition of the term.

Secondly, Garcia and Buchwald

(1964) are attempting to account for the stimulus properties of radiation by pointing out that it leads, as do most other conventional stimuli, to arousal which may be detected electrophysiologically or behaviorally. That radiation leads to EEG changes which resemble those usually associated with sensory induced arousal has been previously suggested by both Soviet (Livonov, 1962) as well as some Western investigators (Gangloff, 1962). Behavioral arousal after low doses, however, was not definitely demonstrated until a year ago (Hunt & Kimeldorf, 1962). It is unfortunate that because of the long publication lag in the *Psychological Bulletin* during the time that my paper was in press a considerable amount of progress had occurred in the area affecting some of the published conclusions. In all fairness, however, if Garcia and Buchwald call attention to their own experiments to account for some of the previously observed stimulus characteristics of radiation, they should mention also some other recent publications from other laboratories which also bear directly upon the problem. Barnes (1962), Hunt and Kimeldorf (1962), Andrews and Peterson (1962), to mention just a few studies, have all published pertinent material.

Finally, to account for the complex

<sup>1</sup> Supported by Public Health Science research career program award No. K3-MH-16, 416-01, National Institute of Mental Health.

and diverse Soviet data on the effect of irradiation in conditioning studies in terms of the short-term arousal effect noted by Garcia and co-workers is much too simplified an explanation. Without going into any great detail I would like to mention just a point or two to illustrate the difficulties encountered in trying to elucidate the role of radiation in conditioning. In all of the radiation-induced spatial avoidance conditioning studies the effects could not be obtained with doses of less than approximately 100 roentgens. Furthermore, the phenomenon was always observed either immediately following irradiation or within a period of less than 36 hours following the conclusion of irradiation. Low doses have no effect and neither do high ones (Andrews & Peterson, 1962). In the Garcia, Buchwald, Feder, and Koelling (1962) study the inhibition of bar pressing occurred in animals which, during the conditioning series, were subjected to an unknown quantity of radiation. (The authors did not state the dose rate during the conditioning series. If the dose rate was the same as during the experiment proper, then the animals must have received 400 roentgens.)

Now in the Soviet studies changes in conditioning were observed after exposures of a few roentgens or less, and these changes sometimes lasted for several days after the cessation of exposure or even for 1 year in one case (Samoylova, 1959). Also, the changes involve different parameters of conditioning in different studies. To state that diverse phenomena observed by Soviet conditioning investigators can be explained on the basis that irradiation affects the arousal mechanism does not carry a sufficient amount of explanatory power to account for the diversity of changes which has been observed.

It should be noted that in my own

review I have stated that the Soviet findings are probably explicable in terms of performance rather than associative changes. As a matter of fact, already in my 1956 review I had noted that radiation affects motivational and emotional variables and that the effects should be most apparent in performance in which the extrinsic incentives are relatively small. This gross generalization is probably still valid. The details in this area of research are rapidly accumulating, and Garcia and co-workers have made here notable contributions.

It is unfortunate that the long publication lag has made my 1963 review paper less up to date than desirable.

#### REFERENCES

- ANDREWS, H. L., & PETERSON, D. C. Variations in radiation recognition by the mouse. *Radiat. Res.*, 1962, 17, 514-520.
- ARBIT, J. Avoidance conditioning through irradiation: A note on physiological mechanisms and psychological implications. *Psychol. Rev.*, 1958, 65, 167-169.
- BARNES, C. D. Peripheral neural paths mediating avoidance of radiation in rats. *Amer. J. Physiol.*, 1962, 203, 379-382.
- FURCHTGOTT, E. Behavioral effects of ionizing radiations. *Psychol. Bull.*, 1956, 53, 321-334.
- FURCHTGOTT, E. Behavioral effects of ionizing radiations: 1955-61. *Psychol. Bull.*, 1963, 60, 157-199.
- GANGLOFF, H. Acute effects of X-irradiation on brain electrical activity. In, *Effects of ionizing radiation on the nervous system*. Vienna, Austria: International Atomic Energy Agency, 1962. Pp. 123-135.
- GARCIA, J., & BUCHWALD, N. A. Note on Furchtgott's "Behavioral effects of ionizing radiations: 1955-61." *Psychol. Bull.*, 1964, 61, 234-235.
- GARCIA, J., BUCHWALD, N. A., FEDER, B. H., & KOELLING, R. A. Immediate detection of X-rays by the rat. *Nature, London*, 1962, 196, 1014-1015.
- GARCIA, J., KIMELDORF, D. J., & HUNT, E. L. The use of ionizing radiation as a motivating stimulus. *Psychol. Rev.*, 1961, 68, 383-395.
- HUNT, E. L., & KIMELDORF, D. J. Evidence for direct stimulation of the mammalian

- nervous system with ionizing radiation. *Science*, 1962, 137, 857-859.
- LIVANOV, M. N. The sensitivity of the nervous system to low-level radiation. In, *Effects of ionizing radiation on the nervous system*. Vienna, Austria: International Atomic Energy Agency, 1962. Pp. 471-484.
- RAZRAN, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, 68, 81-147.
- SAMOYLOVA, L. G. [Concerning certain mechanisms of the influence of low doses of chronic total-body X-irradiation on the higher nervous activity and certain vegetative functions of white rats.] *Med. Radiol.*, 1959, 4(8), 13-17.

(Received May 24, 1963)

# EXACT MULTINOMIAL SIGNIFICANCE TESTS: NOTE ON A PAPER BY ALPHONSE CHAPANIS

N. T. GRIDGEMAN

National Research Council of Canada, Ottawa

Table is given of the 19 possible orderings of 12 items randomly distributed into 3 categories.

Consider two indistinguishable urns, one of which contains equal and the other unequal, numbers of red, white, and blue balls. A sample of 12 is drawn at random, with replacement, from each. Suppose that the results are:

	red	white	blue
Urn A	4	4	4
Urn B	2	3	7

Which urn then is the more likely to house the uniform distribution? Com-

mon sense supplies a ready answer. And Alphonse Chapanis' (1962) paper although primarily concerned, as its title indicates, with single-sample work, has a table (Number 3, page 309) that should furnish a *precise* answer. Surprisingly, this answer favors Urn B. Intuition is sometimes a fallible guide in problems of statistical inference, but this sharp clash between the intuitive and the formal seems worth looking into.

An explanation is to be found in the

TABLE 1

EXACT AND  $\chi^2$  PROBABILITIES OF THE 19 POSSIBLE OUTCOMES WHEN 12 ITEMS ARE RANDOMLY DISTRIBUTED INTO 3 CATEGORIES

Outcome	Number of:		Exact cumulative probability	$\chi^2$ probability	
	Permutations per outcome	Arrangements per permutation		Uncorrected	Corrected for continuity
12 0 0	3	1	.000006	.00001	.00003
11 1 0	6	12	.000141	.00010	.00030
10 2 0	6	66	.000886	.00091	.00104
10 1 1	3	132	.00163	.00117	.00248
9 3 0	6	220	.00412	.00525	.00674
8 4 0	6	495	.00970	.0183	.0267
9 2 1	6	660	.0172	.00866	.0126
7 5 0	6	792	.0261	.0388	.0440
6 6 0	3	924	.0313	.0498	.0725
8 3 1	6	1980	.0537	.0388	.0440
8 2 2	3	2970	.0704	.0498	.0725
7 4 1	6	3960	.115	.106	.135
6 5 1	6	5544	.178	.174	.253
7 3 2	6	7920	.267	.174	.253
6 4 2	6	13860	.424	.368	.417
5 5 2	3	16632	.518	.472	.606
6 3 3	3	18480	.622	.472	.606
5 4 3	6	27720	.935	.779	.883
4 4 4	1	34650	1	1	1
Total	91				

Note.—Totality of arrangements = 531,441.

build-up of Chapanis' Table 3, which gives cumulative probabilities for all possible "outcomes" in 12-item samples from a discrete equipartite trinomial population. These outcomes, 19 in number, are however cumulatively ordered by their *gross* probabilities, whereas the requirement is an ordering by the probabilities of the component permutations. For instance, the outcome 2-3-7 has 6 permutations; and in terms of the urn model they would be: 2 red, 3 white,

7 blue; 2 red, 3 blue, 7 white; 2 white, 3 red, 7 blue; etc.

Table 1 gives the proper ordering. The exact test is of course better than the  $\chi^2$  test, but the discrepancies are now fewer and smaller. And there are no clashes with intuition.

#### REFERENCE

CHAPANIS, A. An exact multinomial one-sample test of significance. *Psychol. Bull.*, 1962, 56, 306-310.

(Received May 15, 1963)

# Psychological Bulletin

## COMPETING THEORIES OF RECEPTOR EXCITATION IN THE RETINA:

### A SYMPOSIUM

HARRY HELSON

*Kansas State University*

Introductory remarks to papers by Enoch, McConnell, Boynton, and Blackwell in a symposium bearing the above title at the 1962 conference of the American Psychological Association in St. Louis.

The following two papers by Enoch and McConnell and the discussions by Boynton and Blackwell constituted a symposium bearing the above title which was held at the Annual Conference of the American Psychological Association in St. Louis, 1962. In spite of their technicality there was an overflow attendance to hear these papers thus attesting the wide interest in the latest work and thinking on mechanisms of color vision. Both the main presentations and the discussions show how far recent visual theory has progressed beyond the classical formulations of Helmholtz, Hering, and Ladd-Franklin which were chiefly designed to correlate phenomenological data with supposed retinal processes. As these papers so well show, theory has advanced to the point where interactions of light and the finer structures of the retina and interactions of

light with photosensitive pigments are now in the center of attention. Furthermore, the adequacy of current theories is judged by reference to quantitative data obtained by meticulous biophysical and biochemical methods as well as by their agreement with psychophysical measurements.

There is much more in the papers by Enoch, McConnell, Boynton, and Blackwell than can be conveyed in a short introduction. Those interested in learning about this area, as well as those committed to either a physical or chemical type of color theory or to some third approach, will find these papers worthy of study. The Editor of the *Bulletin*, who served as Chairman of the symposium, is proud to present material of this caliber and importance in its pages.

(Received March 3, 1963)

# PHYSICAL PROPERTIES OF THE RETINAL RECEPTOR AND RESPONSE OF RETINAL RECEPTORS<sup>1</sup>

JAY M. ENOCH

*Washington University Medical School, St. Louis*

The author reviews studies of the waveguide characteristics of retinal receptors. These properties are present in human photoreceptors. Some of the possible roles played by these physical properties are discussed, with emphasis placed upon color-vision mechanisms at the receptor level. The author also considers approaches toward the determination of the role(s) played by the waveguide properties, by the photosensitive pigments, and by "other" mechanisms in the coding of the signal at the detector level. He describes his initial experiments directed toward the development of an indicator of response (or marker) in single mammalian retinal receptors.

In recent years, this investigator has described the waveguide characteristics of retinal receptors in rats, monkeys, and humans (Enoch, 1960, 1961a, 1961b, 1961c, 1963a). The physical characteristics of the receptors give rise to nonuniform distributions of energy in the photosensitive, pigment-bearing, retinal receptor outer segments. These nonuniform distributions of energy vary as a function of wavelength. That is, each receptor passes a given wavelength band(s) and this energy is distributed spatially in the receptor in certain specific patterns. Since the incident energy is broken up, or separated, into different wavelength bands in the same and neighboring receptors, and since different wavelengths give rise to different chromatic sensations, one may ask if the waveguide characteristics represent the first step in the coding of the chromatic aspect of the visual stimulus. Because of

the considerable complexity of these characteristics,<sup>2</sup> one cannot simply fit these findings into classical theoretical frameworks. At this point, what is needed is not another color-vision theory, or theories; rather, we require a method (or methods) which will enable us to decide which properties define the actual response of the retinal receptor. A crucial question is: How can we precisely determine the role played by waveguide characteristics, the role played by the photochemical properties (Dartnall, 1957; Rushton, 1962; Wald, 1959), and the role(s) played by "other" mechanisms (Ingelstam, 1956),<sup>3</sup> in the response of the receptor, and in the coding of the visual stimulus? The plethora of theories of vision and color vision can only exist in our vacuum of knowledge

<sup>1</sup> This research has been supported in part by Grant B-2168 from the National Institute of Neurological Diseases and Blindness, National Institutes of Health, Public Health Service, Bethesda, Maryland.

This paper was presented as part of the symposium on "Competing Theories of Receptor Excitation in the Retina" at the meeting of the American Psychological Association, September 3, 1962.

<sup>2</sup> While the author may say with confidence that the waveguide form of energy transfer is present in retinal receptors, he cannot yet say that the distribution observed experimentally in any given receptor is of necessity the same as that which was present during life. It is possible that some of the complexities are introduced as a result of the experimental technique. This is another reason for the author's interest in the techniques discussed toward the end of the paper.

<sup>3</sup> This reference is one example of a number of other mechanisms which have been postulated.

concerning the events occurring between the bleaching of the photosensitive pigment and the activation of the neurological chain (and, ultimately, the visual response).

In this paper, the author will review his findings relative to waveguide properties present in retinal receptors of many species (including human); he will discuss the possible role(s) played by these properties in color vision; he will consider some methods of attacking the problem of the response of the single retinal receptor; and he will present samples of his first efforts directed toward the solution of this problem.

#### THE RETINAL RECEPTOR AS A WAVEGUIDE

The term "waveguide" essentially describes the function of the receptor. The luminous energy is literally guided, or channeled, by the receptor into its outer segment, that is, the zone containing the photosensitive pigment. Of significance are the characteristics of retinal receptors acting as waveguides, and the waveguide form of energy transfer or transmission.

Energy, when transferred in a waveguide, is propagated in the form of certain specific waveguide modal patterns. A typical modal pattern photographed at or near the termination of a cone receptor is shown in Figure 1. These patterns, which are the result of interactions of energy in the receptor, are described by a series of summations and cancellations of the energy. They are best observed and photographed in monochromatic light. These results may be predicted to a considerable extent on the basis of mathematical models (Snitzer, 1961; Snitzer & Osterberg, 1961) meeting the conditions of cylindrical dielectric waveguides.

It is important to realize that these patterns extend slightly beyond the cell



FIG. 1. A typical retinal receptor waveguide modal pattern as photographed in monochromatic light— $\lambda = 580$  millimicrons. (The microscope was focused upon the termination of the retinal receptor, the incident energy subtended a cone angle at the retina of 3 degrees, and the receptor was well oriented. The energy passed through a retinal cone near the central fovea of a squirrel monkey. A length of 1 micron is indicated.)

boundary. If the separations between the receptors are not sufficient, this energy may leak into surrounding cells. This could reduce visual resolution and influence transmission of energy in neighboring cells. At this time, it is not believed that this interreceptor leak factor is too significant in the normal (human) retina. However, the author reserves judgment on this point until further studies are completed.

One observes these patterns readily if some care is taken in preparing the specimen. The instrument required is essentially a schematic eye. One focuses an intense source of light upon a piece of freshly dissected retina oriented as it might be in the eye. It is important to limit the angular subtense of the incident beam to angles normally subtended by the pupil of the eye. The dissected specimen, which is placed in a special chamber to avoid crushing, is observed with a microscope. One focuses the microscope at or near the tips of the photoreceptor outer segments and observes modal patterns emerging from the cells after that energy has passed through the cells. Samples of typical human retinal receptors are shown in Figure 2.

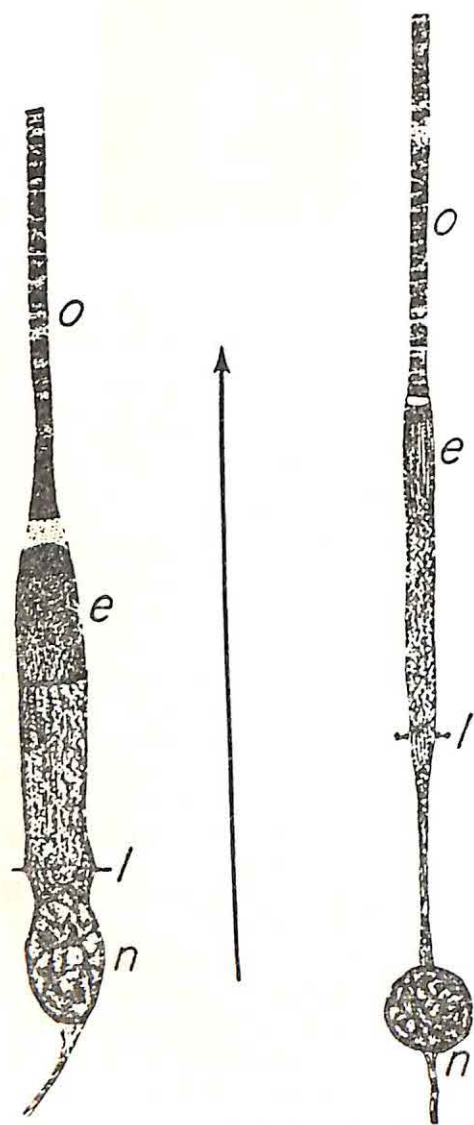


FIG. 2. Hypothetical human retinal receptors. (The arrow indicates the direction normally traversed by the incident light. The cell to the left is a parafoveal cone, and the receptor to the right is a rod. The parts of the cell are: o, the pigment-bearing outer segment; e, the ellipsoid which contains densely-packed oriented mitochondria; l, the external limiting membrane; n, the cell nucleus. The synaptic region is not shown. Courtesy of G. Walls, *The Vertebrate Eye*. Bloomfield Hills, Michigan, Cranbrook Press, 1942.)

Those experimental factors which vary the patterns propagated are the wavelength of the light, the angle of incidence of the incident light, and the plane of focus of the microscope. Polarization of the incident light has little apparent effect upon the result.

Perhaps the best way to summarize the phenomena one observes is to describe the events which take place when one introduces certain variables. First, let us consider an experiment where a non-foveal retinal sample is illuminated by a perpendicularly incident, narrow wavelength band, beam of light. The light passes through the retina as it might in the living eye. One racks the microscope down from a plane of focus above and behind the retina. The first entities one observes through the microscope are the remnants of the pigment epithelial layer. The pigment granules are clearly visible, particularly in monkeys. Passing down past this level, one focuses at, or near, the terminations of the rod receptors. Here one sees a modal pattern in each and every receptor. The types of patterns observed in reasonably well-oriented receptors fall into certain limited distinct categories. In some instances one sees interactive forms. That is, if physical conditions are such that more than one modal pattern may propagate (that is, pass on down the receptor), and more than one pattern does propagate, then these patterns may interfere with each other, and one observes the resultant forms. In these cases, if the two distributions are proceeding with differing velocities down the receptor, one may anticipate the presence of essential standing-wave patterns with concentrations of energy at certain points along the receptor outer segment. These points of concentration indicate where the two patterns are "in step." Schroeder (1961) of the Radio Corporation of America and Snitzer and Polanyi (1961) of

American Optical Company have considered possible interactions.

Between the receptors one often sees microfibrils (penetrating for a short distance) from the pigment epithelial layer. These microfibrils contain pigment granules. In some receptors the modal patterns are much more intense than in others. There are two essential reasons for this, either the receptor is less well oriented, or it is not transmitting much energy of this particular wavelength. The differentiation between these two possibilities is made by changing the plane of focus and/or the wavelength of the incident light. It is also of interest that, when focusing on the rod terminations, there are reasonably regularly disposed darkened areas. These turn out to be the locations of the cones when one continues to rack the microscope in the same direction.

When one changes the plane of focus in this manner, the patterns usually become less bright than at the terminations, their brightness fluctuates, and the modal patterns may change. The important point is that some pattern is seen at almost all planes of focus in the receptor. This energy probably represents that which is leaking out of the receptors prior to reaching the receptor terminations. At a certain point one sees the highly intense cone terminations. They are more intense than the rod terminations because of the essentially integrative action of the ellipsoid portions of these cells which collect much of the energy from the larger diameter cone inner segments and guide that energy into the smaller diameter outer segments. The more oblique the angle of the incident light, the smaller is the amount of energy which reaches the outer segment. When one continues to focus the microscope down to still deeper levels, at the point where the cone diameters enlarge, one often observes one or more modal-pattern

changes. This is one factor which suggests that the final transmission characteristic of the receptor is determined in the outer segment itself.

If on varying the plane of focus of the microscope the observed pattern(s) appear to shift sidewise in position in the field, this indicates a defect in receptor alignment in the specimen. The magnitude and direction may be measured. Few retinal receptors are perfectly coaxial with the microscope—in many cases this may be an artifact associated with the preparation.

If one continues to rack the microscope down after passing into the region of the receptor inner segments, at some point the light breaks up into a myriad of small points. Presumably one has passed the external limiting membrane. Backing the microscope off all the way in the other direction, one observes the rapid dissipation of the radiating patterns after passing the receptor terminations.

If one now refocuses the microscope at, or near, the receptor terminations and varies the wavelength (or color) of the incident light, one observes sudden changes in modal pattern as well as marked variation in relative transmission of energy. Similarly, if one varies the obliquity of incidence of the incident light, one also sees sudden shifts in modal pattern. Increasing wavelength and increasing obliquity of angle of incidence often (but not always) produce similar changes in modal pattern at the same receptor cell outer segment termination.

When one introduces obliquity of incidence, one observes a decrease in the amount of energy passing all the way through the receptors. As the angle of obliquity increases, the decreases in apparent brightness in cones is somewhat more rapid than in rods. On the basis of innumerable observations, the author

feels that the individual rods and cones are quite a bit more directionally sensitive than psychophysical functions suggest. The latter functions probably should be considered as summed group effects.

If white light is used instead of monochromatic light, many interesting events are observed. First, the retina takes on a richly variegated, multihued, mosaic-like appearance. This effect is related both to differences in transmissivity and to the changes in modal pattern as a function of wavelength. It is clear that a physical separation of the incident wavelengths takes place. If one changes the angle of incidence, the photoreceptor transmissive characteristics change as a function of wavelength (in addition to other changes in modal patterns and total transmissivity). If one changes plane of focus, one sees different wavelength components leaking out of the retinal receptor at different levels.

In summarizing these events, one may say that the incident energy is broken up in the retinal receptors. Different wavelengths are located or concentrated at different spatial positions in different receptors and within the same receptor. These nonuniform distributions represent the input stimulus distributions to the individual retinal receptors. A receptor can only respond to, and hence detect, that which is delivered to its photosensitive pigment(s).

#### ROLE(S) PLAYED BY THESE PROPERTIES IN COLOR VISION

The challenging question which must now be considered is: What functional role does this form of energy transfer play in the visual process? The most obvious consideration, and the one we shall discuss here, is that of color vision. The color-vision problem at this level is one of coding the information present in the image by the retinal receptor. Are

waveguide properties used by the visual system for coding purposes?

We do know that at some point in the system a given luminous stimulus may be described in terms of three variables. However, this does not mean that the response is limited to this number of variables at the receptor. It has generally been held that there are three or more different photosensitive pigments in cones, and that they provide the necessary wavelength separation for coding purposes. The research reported in this paper demonstrates that wavelength separation does, in fact, take place physically in the receptors. That is, the light is broken up into different wavelength components located at different places in the same, and in different cells. If this information is used by the visual system, the manner in which it is used is dependent upon the number of different photosensitive pigments present and how the photosensitive pigment molecules are distributed. With the volume of information present in the physical distribution, it is immediately clear that only one type of photosensitive pigment is required by the visual system. However, this does not necessarily mean that only one type of retinal cone pigment is present.

Following this line of reasoning, one may ask several questions. What part of the coding task takes place in a single receptor cell? Considering the transmission characteristics of these cells, is it possible that one cell can accurately describe the entire input distribution to the nervous system? Is the response of any single cell ever independent (Sjostrand, 1958)? How well does the distribution of photosensitive pigment correlate with the physical distribution of the energy?

The separation of the incoming energy into various wavelength (color) components occurs in at least five differ-

ent ways. First the waveguide modal pattern often changes in a single cell with variation of wavelength. This means that the transmitted energy is redistributed in the outer segment of the cell. In order for this distribution of energy to be used for coding purposes, it is clear that the cell, acting as a detector, must be able to determine that a change has occurred. This information could be utilized if the photosensitive pigment or pigments were nonuniformly distributed in the cell. As an alternative, the response of the cell might be dependent upon the location of the bleached pigment in the cell, for example, distance from the surface membrane, etc. Clearly, if such mechanisms exist, changes in physical characteristics of the cell could lead to visual anomalies through poor correlation of the concentrations of light energy and the sites of specific photosensitivity. Also implied, in a mechanism such as this, is the ability of the cell to transmit or to transfer centrally the information acquired. How much of the entire coding task is achieved in a single detector is dependent upon its complexity (or sophistication, or intelligence) as a detector and its independence of action.

A second mechanism for separating the incident light into separate wavelengths is mode interaction. The standing waves set up in this manner result in local concentrations of energy along some cone outer segments. Their loci vary with wavelength. The only difference between this case and the first case is that the concentrations of energy are probably not so discretely placed in the receptor outer segment. The conditions for utilization of that information are precisely the same.

The third mechanism by which the input distribution of energy is separated is related to the selective manner by which energy is transmitted in the re-

ceptor. When white light is focused upon the receptor and one looks at the terminations of the receptors, some appear red, others blue, etc. (Enoch, 1961a). Thus, different receptors selectively transmit energy distributions of different wavelengths more or less efficiently. This is a filterlike effect. Snitzer and Polanyi (1961) refer to this as "flash-point transmission." The effect is sensitive to angle of incidence of the energy at the retina; and, hence, the receptor cannot be considered a true filter. This filterlike action requires only a relatively unsophisticated detector for its utilization. The main requirement is that the detector be "aware" that its physical characteristics favor the transmission of a given wavelength band. Whether or not this information is actually used by the system, this phenomenon must bias the result. Again, it must be determined how well correlated are the transmissive properties and the detecting properties.

A fourth mechanism arises out of the third one. If white light is incident at the receptor, and red light is seen emerging from the receptor termination, one may inquire as to what happened to the remaining energy. Obviously, some is reflected or scattered, and does not propagate, or is not transmitted in the receptor. Other energy which is transmitted leaks out at different points prior to the receptor termination. That is, different levels within the receptor outer segment probably have different wavelength compositions. In order to use this information one must again consider the sophistication of the detector, and the response transfer mechanisms. In our experimental preparations most of the photosensitive pigment was bleached; hence, there was little absorption.

Lastly, the energy propagating outside the cell does not come to bear upon the photosensitive pigment. The magnitude

of this component varies with wavelength and modal pattern. The result is a filterlike effect which almost certainly biases the result to some extent. Myers (1962) recently considered this phase of the problem.

One notes that all of these effects are present in both rods and cones. Because vision dominated by rod receptor response apparently has no chromatic component, one cannot say that waveguide mechanisms do not play a role in color vision. One may argue simply that rods do not have any mechanism present which is capable of utilizing the information present. By the same token, one cannot claim other than a biasing role for waveguide characteristics in cone vision because wavelength separation is present. Such bias could lead to an anomalous visual response. It becomes necessary now to design and conduct experiments which will indicate the true role of these phenomena in vision.

In some well-oriented rod-free central foveal areas, one notes that the transmissive properties of different cells are apparently not randomly distributed. Rather, at regular intervals a transmissive pattern appears to be repeated. This is particularly evident when monochromatic light is employed. One can argue that such patterning of these properties indicates probable utilization of these characteristics.

#### DEFINING THE RESPONSE CHARACTERISTICS OF THE SINGLE RECEPTOR CELL

In order to evaluate the role of waveguide characteristics, photochemical properties, and other possible mechanisms in receptor response, it is desirable to study the parameters defining the response of single receptors. The waveguide characteristics of a given receptor are dependent upon the configuration of the cell, the diameters of its component

parts, the index of refraction of its parts relative to the surrounding medium, the separation of the cell from its neighbors, and its orientation relative to the optical system of the eye. The presence of, and distribution of a photosensitive pigment or pigments within single or different cells probably, but not necessarily (Brindley, 1960; Ingelstam, 1956), defines the transducer.

It is important to realize that if a photosensitive pigment having peak sensitivity in the red is measured, this fact does not prove that the system is aware that this is an effective red detector. Similarly, predominantly red transmission due to waveguide characteristics, or the presence of a red oil droplet which acts as a filter in the inner segment (e.g., in the pigeon) does not define the coding mechanism. One must consider these factors in terms of probability statements. One must seek to correlate response at a level (or levels) central to the outer segment with those physical, morphological, and chemical factors concerning the detector which one can determine. Presumably, those physical and/or chemical characteristics which have the highest correlation with response (in the normal) define the coding mechanism. The solution of this problem will not be simple.

The bleaching of the photosensitive pigment bears a functional relationship to response, but the bleaching of any given photosensitive pigment molecule may not result in a response. Some of the mechanisms (relating to the possible role(s) played by waveguide properties) suggested above depend on this possibility. This is not a new concept. For example, Wald (1954) in his compartment theory makes use of this point.

In all of our discussions of cell response we have tended (largely through lack of knowledge) to overlook the role of the secondary mediators of receptor response. Since the time of Hecht, we

have realized that the bleaching of the photosensitive pigment only "triggers" retinal response. The response characteristics of the receptor may be determined to a considerable degree by the properties of the succeeding chemical reactions and/or physical events. For example, if receptor-cell excitation is mediated through a chemical reaction(s) induced by the breakdown of the photosensitive pigment, we have no right to assume the constituent substances are in infinite supply, cannot be (partially, or locally, or largely, or temporarily) depleted, and are uniformly distributed throughout the receptor outer segment or other critical zone.

Microspectrophotometry can tell us the nature of the photosensitive pigments present (and probably how much of a given pigment has been bleached), and this author's methods will ultimately provide information relative to waveguide properties. We now need to turn our attention toward the definition of indicators or markers of individual receptor response. We must know which receptors have been actively responding when excited by a given stimulus, and in turn, define their properties. It will also be important to see how these properties differ from those present in nonresponding cells.

Ideally, these indicators must represent activity induced by the bleaching of the photosensitive pigment. Activity at the bipolar level is probably too late in terms of defining initial response characteristics because of the problem of localizing which receptor(s) induced a given bipolar cell to respond, and because several receptors may participate jointly in inducing bipolar activity through complex synaptic junctions. Even the study of the synaptic zone in the receptor may not prove of value. Sjostrand (1958) has demonstrated the presence of synaptic connections between

receptor terminations in the guinea pig. Thus, we must seek indicators or markers of response in the photoreceptor itself, central to the outer segment (so that the indicator only minimally interferes with the evaluation of the properties of the outer segment, and so that we know that sequential activity has been induced), but peripheral relative to the synaptic zone. One may approach this question through electrophysiology, or histochemistry, or biochemistry, or ultrastructural research.

Hagins, Zonana, and Adams (1962) have been able to detect a receptor response electrophysiologically in single squid receptors. The morphology of the squid retinal receptor is somewhat different from that in higher species. This author hopes that Hagins and his co-workers can extend their techniques to other species. There have been several studies of the histochemistry of rod and cone receptors in mammals including humans in recent years (Berkow & Patz, 1961; Cogan & Kuwabara, 1959, 1960; Kuwabara & Cogan, 1959, 1960; Pearse, 1961). Pearse (1961) has been able to detect a difference in staining between a dark-adapted and light-adapted rat retina in the pigment epithelial layer. A very important approach using biochemistry is that employed by Lowry, Roberts, and Lewis (1956). Various other biochemical studies relating to the problem have been conducted (Hanawa & Kuge, 1961; Pirie & Van Heyningen, 1956). Lastly, one may turn to electron microscopy to evaluate changes due to receptor response. Fernandez-Moran (1961) has recently reported some interesting work in the outer segment.

When evaluating receptor response it is first necessary to determine techniques indicating physical or chemical changes-of-state in the receptor due to a given luminous stimulus. The change of state must be validated (Enoch, 1963b). That

is, an effort must be made to relate the change of state to the response system through correlation of its response characteristics, and, ultimately, through the determination of the actual physical or chemical relationships. Lastly, one must determine which characteristics of the cell (waveguide, photochemical, or other) correlate with the change of state for a given set of stimulus conditions. These ends must be achieved with minimum disturbance of physiological conditions.

#### PRELIMINARY INVESTIGATIONS OF RECEPTOR RESPONSE

As a first step in the study of receptor response this author turned toward histochemistry. His first paper on the subject has recently appeared (Enoch, 1963b). The following paragraphs summarize that work briefly.

The author has chosen the ellipsoid of mammalian retinal receptors for study, and he has been able to demonstrate a difference in staining in light and dark-adapted eyes (Figures 3 and 4).<sup>4</sup> The light-adapted retina may be exposed to the light either after being excised, or during life, prior to dissection. The rat retinas shown in the figures were dark adapted for 15 hours prior to dissection, both were dissected in dim red light, and both were incubated in a special medium in the dark. The light-adapted retina was exposed to the stimulus for 10 minutes starting about 15 minutes prior to dissection. The luminance level of the extended stimulus evaluated in the plane of the entrance pupil of the rat was approximately 10,000 millilamberts (human photopic evaluation). The source of light was a tungsten filament with heat and neutral density filters inter-

<sup>4</sup> The author wishes to acknowledge the generous advice and encouragement given to him during this work by members of his department, particularly by Marguerite Constant and Bernard Becker.

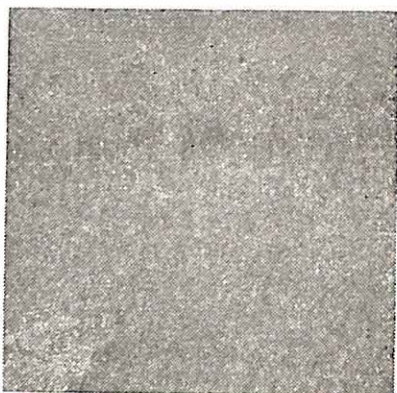


FIG. 3. This photomicrograph is an example of a dark-adapted rat retina which has been incubated in a medium containing a tissue-culture medium (TC 199), succinate, and nitro-blue tetrazolium. (No stain was observed in this section.)

posed. The pupil was dilated. Similar experiments have been conducted on rabbits and squirrel monkeys. The effect is present in both rods and cones. The eye may be exposed to light a significant period of time prior to dissection and the difference in staining still obtained. This allows regeneration of a substantial portion of the photosensitive pigments.

These experiments have only dealt

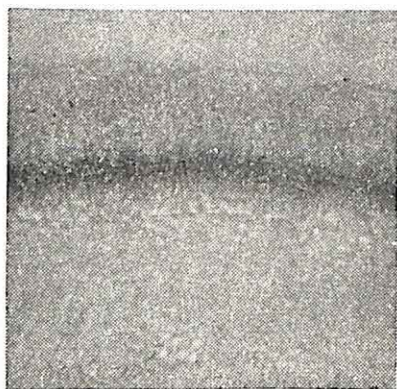


FIG. 4. The second retina of the same rat. (The eye was light adapted, after having been dark adapted, shortly before dissection of the retina from the living animal. The retina was incubated in the same manner as the specimen shown in Figure 3. The ellipsoids—see Figure 2—of these rod receptors stained heavily.)

with the first aspect of the problem, that is, the determination of a technique indicating a change of state in a receptor due to a given luminous stimulus. The technique has not yet been fully validated. The author hopes that this reaction will prove to be the indicator which he seeks. Obviously, this step only represents the beginning of a rather complex problem. If the validity of this indicator can be established, it may be possible to significantly advance our knowledge of coding in the retinal receptor.

# REFERENCES

- BERKOW, J., & PATZ, A. Histochemistry of the retina. *AMA Arch. Ophthalmol.*, 1961, 65, 820-831.
- BRINDLEY, G. S. *Physiology of the retina and visual pathway*, London, England: Edward Arnold, 1960.
- COGAN, D. G., & KUWABARA, T. Tetrazolium studies on the retina. *J. Histochem. Cytochem.*, 1950, 7, 334-341.
- COGAN, D. G., & KUWABARA, T. Tetrazolium studies on the retina. *J. Histochem. Cytochem.*, 1960, 8, 380-384.
- DARTNALL, H. J. A. *The visual pigments*. London: Methuen, 1957.
- ENOCH, J. M. Waveguide modes: Are they present, and what is their possible role in the visual mechanism? *J. Opt. Soc. Amer.*, 1960, 50, 1025-1026.
- ENOCH, J. M. Nature of the transmission of energy in the retinal receptors. *J. Opt. Soc. Amer.*, 1961, 51, 1122-1126. (a)
- ENOCH, J. M. Visualization of waveguide modes in retinal receptors. *Amer. J. Ophthalmol.*, 1961, 51, 1107-1118. (b)
- ENOCH, J. M. Waveguide modes in retinal receptors. *Science*, 1961, 133, 1353-1354. (c)
- ENOCH, J. M. Optical properties of the retinal receptors. *J. Opt. Soc. Amer.*, 1963, 53, 71-85. (a)
- ENOCH, J. M. The use of tetrazolium to distinguish between retinal receptors exposed and not exposed to light. *Invest. Ophthalmol.*, 1963, 2, 16-23. (b)
- FERNANDEZ-MORAN, H. The fine structure of vertebrate and invertebrate photoreceptors as revealed by low-temperature electron microscopy. In G. Smelser (Ed.), *The structure of the eye*. New York: Academic Press, 1961. Pp. 521-556.
- HAGINS, W. A. ZONANA, H. V., & ADAMS, R. G. Local membrane current in the outer segments of squid photoreceptors. *Nature*, 1962, 194, 884-847.
- HANAWA, I., & KUGE, K. The effect of light intensity upon the oxygen consumption of the isolated outersegments of rods. *Jap. J. Physiol.*, 1961, 11, 38-43.
- INGELSTAM, E. Electromagnetic theory of the primary mechanism of colour vision. In, *Problems of contemporary optics*. Florence, Italy: Istituto Nazionale di Ottica, 1956. Pp. 640-668.
- KUWABARA, T., & COGAN, D. Tetrazolium studies on the retina. *J. Histochem. Cytochem.*, 1959, 7, 329-333.
- KUWABARA, T., & COGAN, D. Tetrazolium studies on the retina. *J. Histochem. Cytochem.*, 1960, 8, 214-224.
- LOWRY, O. H., ROBERTS, N. R., & LEWIS, C. The quantitative histochemistry of the retina. *J. Biol. Chem.*, 1956, 220, 879-892.
- MYERS, O. E. Spectral sensitivity of visual receptor cells. *Nature*, 1962, 193, 449-451.
- PEARSE, A. Localization of oxidative enzymes in rat and chick retina in various physiological conditions. In G. Smelser (Ed.), *The structure of the eye*. New York: Academic Press, 1961. Pp. 53-72.
- PIRIE, A., & VAN HEYNINGEN, R. *Biochemistry of the eye*. Oxford, England: Blackwell Scientific Publications, 1956.
- RUSHTON, W. A. H. *Visual pigments in man*. Liverpool, England: Liverpool Univer. Press, 1962.
- SCHROEDER, A. C. Waveguides and the receptor mechanism in color vision. *J. Opt. Soc. Amer.*, 1961, 51, 909.
- SJOSTRAND, F. S. Ultrastructure of retinal rod synapses of the guinea pig eye. *J. ultrastruct. Res.*, 1958, 2, 122-170.
- SNITZER, E. Cylindrical dielectric waveguide modes. *J. Opt. Soc. Amer.*, 1961, 51, 491-498.
- SNITZER, E., & OSTERBERG, H. Observed dielectric waveguide modes in the visible spectrum. *J. Opt. Soc. Amer.*, 1961, 51, 499-505.
- SNITZER, E., & POLANYI, M. Possible waveguide mode effects in retinal receptors. *J. Opt. Soc. Amer.*, 1961, 51, 476.
- WALD, G. Mechanism of the visual threshold and visual adaptation. *Science*, 1954, 119, 887-892.
- WALD, G. The photoreceptor process in vision. In J. Field (Ed.), *Handbook of physiology*. Vol. 1. *Neurophysiology*. Washington, D.C.: American Physiological Society, 1959. Pp. 671-692.

(Received March 3, 1963)

## CHEMICAL THEORIES

DAVID G. McCONNELL

*Institute for Research in Vision, Ohio State University*

The chemical theories of the visual process advanced by Wald, Dartnall, and others are compared. Their adequacy for scotopic vision is substantial, while incomplete. Almost no theory successfully explains photopic vision. Evidence that rhodopsin is a light-sensitive enzyme is discussed, and it is conjectured that present extraction techniques do not permit comparison of in situ and in vitro data, particularly, in the photopic case,

### BRIEF HISTORICAL BACKGROUND

The history of research in vertebrate vision chemistry is about a century old. Heinrich Müller (1851) first described the reddish color of the outer segments of the retinal rods. In 1876 Boll noted that the color disappeared rapidly after removal of the retina, and before removal if the animal were exposed to bright light. Kühne (1878) then demonstrated that the purplish-red pigment *Sehpurpur* or "rhodopsin" could be extracted in a solution of bile salt, and that the color was stable in the dark but bleached in the light. Köttgen and Abelsdorff (1896) plotted the differences in absorption spectra of bleached and unbleached solutions of the extract for several species, and found absorption maxima for mammals, birds, and amphibians at about 500 millimicrons and at about 540 for fishes. The corresponding pigments are conventionally called "rhodopsin" and "porphyropsin."

Trendelenburg (1904) used a Nernst lamp to bleach solutions of visual purple and measured the efficiency of different parts of the spectrum in bleaching. Though he did not allow for differences in energy of the bleaching lights, he used the same lamp to determine dark-adapted visual sensitivity, and found that it compared closely with the bleaching efficiency of the lamp.

Studies made by Selig Hecht (1920, 1921, 1924), using white light and solutions containing various ratios of

bleached and unbleached rhodopsin, enabled Hecht to determine that the course of bleaching could be described by a monomolecular equation.

Fridericia and Holm (1925) and Tansley (1931) found that vitamin-A-starved rats synthesized visual purple less rapidly than normals, and George Wald (1933) reported direct evidence of the presence of vitamin A in eye tissues. In experiments with frog retina he found a connection between vitamin A and visual purple. He also related to their presence a new substance which he called retinene, extracted from solutions of light-bleached rhodopsin. It remained for Morton and Goodwin (1944) to prove that this pale yellow carotenoid was the aldehyde of vitamin A. Even before then, however, Wald (1939) published evidence that the pigment of fish retinas (porphyropsin) was related to a retinene and a vitamin A different from those in the frog. They were named vitamin A<sub>2</sub> and retinene<sub>2</sub>, and their absorption maxima ( $\lambda_{\max}$ ) are both displaced towards the red about as much as porphyropsin is displaced with respect to rhodopsin. As recently as 1953, Wald argued that only these two visual pigments existed in the rods of vertebrates; this has become known as the rhodopsin-porphyropsin theory.

### THE WALD MODEL

The top structure in Figure 1 is the retinene<sub>1</sub> molecule. Its alcohol is vita-

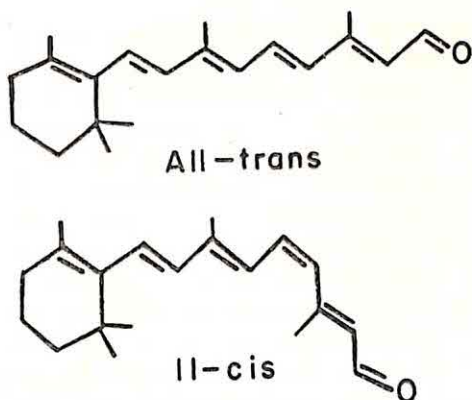


FIG. 1. *All-trans* and *11-cis* (*neo-b*) isomers of vitamin A aldehyde, or retinene.

min A<sub>1</sub>. The retinene<sub>2</sub> structure, and its vitamin A<sub>2</sub> alcohol, have an additional double bond in the ring, accounting for displacement of their absorption spectra to the red. The simplest model for the chemical process underlying vision at the time of Wald's discovery of a relationship between rhodopsin, vitamin A, and retinene involved fragmenting of the retinene molecule from the protein group of rhodopsin, under bleaching by light. Once split off, the retinene was reduced to its alcohol, and the vitamin A was then somehow reincorporated with the protein into the light-sensitive rhodopsin complex. This model had a substantial appeal, because it equated the visual stimulus to bleaching, and the big remaining problem appeared to be how to get the disjoined chromophore back into union with the protein. Things stayed this simple only momentarily, however, because Wald and his co-workers discovered that the bleaching product was the *all-trans* isomer of retinene, while the isomer which in proper solution in the dark would recombine with the protein to yield rhodopsin was the bottom structure in Figure 1—*11-cis* or *neo-b* retinene. Thus, though a pathway was found for regeneration of rhodopsin from solutions of protein and retinene, it was not the same retinene as the photoproduct.

Hence the next problem was to determine how to get *11-cis* retinene from *all-trans* retinene. These were not the only isomers of retinene by any means, but happen to be central to the theory.

The explanation provided by Wald for the difference in the bleaching product and recombining molecule was that the light-sensitive rhodopsin contains the *cis*-isomer only. This is bleached to the *trans*-isomer. Although controversy still exists as to the chemical nature of these products, there is fair agreement as to the sequence. The primary photochemical reaction converts rhodopsin to the intermediate *lumi*-rhodopsin, which is stable only at very low temperatures. At physiological temperature it is thermally converted to *meta*-rhodopsin, which Dartnall (1957) calls transient orange. Both Dartnall, representing the Liverpool school of thought, and the Harvard group under Wald's leadership appear to agree that no isomerization occurs in the primary reaction. Whether isomerization occurs thermally in conversion of *lumi*- to *meta*-rhodopsin apparently depends both on the species of the animal and the orientation of the theorist. The Liverpool scheme involves conversion of transient orange to indicator yellow, so-called because it changes color with pH. The Harvard school disputed this, believing that retinene was produced directly from *meta*-rhodopsin by hydrolysis. The problem hinged upon whether retinene was actually freed off from the protein, as the Harvard group believed, or remained linked to the protein throughout the visual cycle, as the Liverpool group believed. Of equal and related importance was whether the linkage between the protein and the chromophore was sulfhydryl, as the Harvard group believed, or a Schiff base—formed by bonding of the aldehyde retinene with a protein amino group, as favored by the Liverpool school.

In this complicated schema, we have already remarked on the uncertainty of where isomerization occurs. This is intrinsic to the question of at what stage in this model, if any, does the visual process get going? Because they argued that attachment of retinene to opsin still prevailed in the *lumi*- and *meta*-stages, but no longer did after hydrolysis, Hubbard and Kropf (1958a) in remarking on this problem have stated that visual excitation probably precedes the release of retinene and depends upon the light reaction itself, or, at the latest, on the conversion of *lumi*- to *meta*-rhodopsin. However Hagins (1956) showed that bleaching is fast enough to precede the electroretinogram, and for this and other reasons it must be acknowledged that the data are still too incomplete to permit pinpointing the time of origin of the visual sensation. It seems possible, however, as will be discussed in more detail below, that excitation is synonymous with light-induced change in the activity of an adenosine triphosphatase enzyme which the author believes to be one and the same with opsin. Once again, however, the time scale of this change on the order of milliseconds as required to describe excitation is presently lacking.

At the heart of any ultimately accepted chemical theory of visual excitation will certainly be an explanation of the mechanism for regeneration of rhodopsin from the bleached product. While in solution, Hecht showed, bleaching continues to completion and is dependent upon duration and intensity of the light, in the intact retina bleaching is no longer proportional to intensity. Hagins (1955) showed this when he demonstrated that for very short flashes of light, about half of the rhodopsin available *in vivo* is bleached regardless of intensity. Some process which is not understood opposes the bleaching of rhodopsin in the living retina. That this process is concerned with the pigment epithelium has been

suggested in several quarters. Indeed, Kühne had determined as early as 1878 that contact with the pigment epithelium was necessary for the retina to regain its color in the dark, after bleaching. Hubbard (1956) found that an extract of the epithelial pigment which she called retinene isomerase was effective in converting the *all-trans* retinene to the *11-cis* isomer, and that this property was significantly greater in the light than in the dark. Rushton (1958a) argued against the significance of this phenomenon for the *in vivo* retina, because of the results of an experiment he reported from his laboratory. Blue and yellow lights, matched for equal brightness by a dark-adapted human subject, were equally effective in bleaching rhodopsin, as measured by Rushton's apparatus for analyzing light reflected from the live retina. Regeneration following removal of the two lights was also identical. Since retinene in solution absorbs blue over 10 times more than yellow, the blue bleaching should have been more overcome by isomerization than the yellow. Rushton's own model (1958b) entails an equilibrium between retinene and vitamin A, under control of an enzyme-coenzyme complex. In this regard, his model is very close to that of the Harvard group. Since regeneration of the bleached photopigment is intimately tied to the presence of the epithelial pigment, the concept of a diffusion process between it and the receptors has been entertained by the Harvard group. Though such a process may exist, it is not necessarily the process which opposes bleaching. Another interesting possibility is that a very rapidly formed free radical, such as is known by electron spin resonance to be associated with the mammalian retinal melanin pigment (Cope, Sever, & Polis, 1963) opposes bleaching. The possibility of the movement of charge carriers between the receptors and the epithelial pigment

cannot be discounted. Finally, work in the author's laboratory has shown that thoroughly bleached rhodopsin, as well as extracts of bleached retinas, still contain a strong light-sensitive enzyme which recovers substantial activity when left in the dark. It is not necessary for the rhodopsin to recover its typical absorption spectrum in order for the enzyme to recover activity, though more is recovered when retinene is available.

A recapitulation of the Wald model appears in Figure 2. Rhodopsin is converted by light, in the first step, to orange intermediates which are hydrolyzed at sufficiently high temperatures to *all-trans* retinene and its protein *scotopsin*. In the upper of the two horizontal reactions, *trans-cis* isomerization of the retinene occurs, under the control of retinene isomerase and light. In the lower reaction the *trans*-vitamin A is isomerized. The oxidation-reduction exchanges of hydrogens between retinene and the alcohol are controlled by alcohol dehydrogenase and the coenzyme DPN. Ultimately, the *11-cis* isomer of retinene combines in the dark with opsin to resynthesize rhodopsin as fast as it is broken down by bleaching.

Comparison of the density spectrum of human rhodopsin with spectral sensitivity for the dark-adapted human eye has been carefully made by Crescitelli and Dartnall (1953), and exhibits a rather good correspondence. This agreement, and the absence of other photopigments from solutions extracted in the dark from rod outer segments, has made

it very difficult to suppose that another molecule beside rhodopsin is responsible for human scotopic vision. For human photopic vision, however, no model is available which even approaches coherence between chemical and psychophysiological facts.

#### HOW MANY VISUAL PIGMENTS?

In the chicken retina, Wald (1937) was able to extract along with rhodopsin a very small amount of a pigment with  $\lambda_{\max}$  of 562 millimicrons. On the basis of later experiments Wald, Brown, and Smith (1952, 1953) concluded that this pigment—iodopsin—consists of the same retinene *cis*-isomer as rhodopsin contains, but, instead of the rod protein *scotopsin*, a cone protein which they called *photopsin*. Furthermore, just as scotopsin will combine either with the retinene<sub>1</sub> *cis*-isomer to make rhodopsin or with the retinene<sub>2</sub> *cis*-isomer to yield porphyropsin, photopsin should combine with either the retinene<sub>1</sub> or retinene<sub>2</sub> *cis*-isomers to give iodopsin or "cyanopsin," respectively. Cyanopsin, then, is the fish analogue of the chicken cone pigment, iodopsin. In this scheme, there is one rod pigment and one cone pigment in any animal, though pigment pairs may differ from species to species. Crescitelli (1958) has published data on the distribution of the  $\lambda_{\max}$  of bleachable pigments from the retinas of a number of vertebrates. There appears to be very little evidence from his study that interspecies differences can be categorized according to the Wald rhodopsin-porphyrpsin theory. Nor does it appear that in any animal a cone pigment and a rod pigment can be readily differentiated by easily applied extraction techniques. Wald's categorization of the bleachable pigments has been criticized for its failure to apply rigorous tests of the homogeneity of the pigments. In some cases, homogeneity can be argued

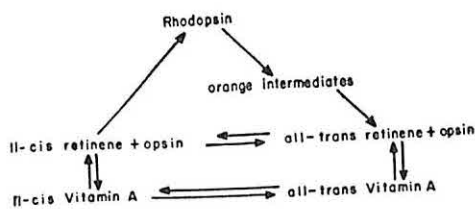


FIG. 2. Simplified diagram of the Wald model of the visual cycle.

from the physical chemistry of the bleaching curves, after extraction. In the case of the cone pigments, however, miniscule quantities are probably extracted, if at all, and they are easily confused with impurities. Moreover, it is not at all certain that all retinal photopigments bleach in the light, even though they may be photosensitive. Nonetheless, most investigators are in agreement with an important (though certainly informal) generalization of Wald's rhodopsin-porphyrpsin theory—that all visual pigments are members of some homologous chemical series. We may recall that porphyropsin, to illustrate, differs from rhodopsin merely in the presence of a single additional double bond in the ring. Extension of the conjugation in this or some other manner is expected to displace absorption toward the red. On the other hand, interruption of the conjugated system in some manner may be expected to shorten the wavelength of the chromophore. Dartnall (1952) pointed out that when the absorption spectra of a large number of visual pigments are slid along the horizontal axis (wavelength) so that their absorption maxima coincide, there is remarkable agreement among them. Yet these considerations alone hardly constitute a sufficient basis for acceptance of the homologous series generalization. Perhaps the greatest arguments for such a theory are parsimony and the existence of analogues in other ubiquitous biological systems such as the porphins.

### NEW APPROACHES

Refinements in knowledge and control of biological chemical systems frequently depend in part on *in vitro* analysis of extracted material such as we have just described in the case of rhodopsin, and in part on synthetic models. Thus far this has been true with the visual pigments and promises to be more so in the

future. Organic synthesis played a major rôle, often ignored, in Wald's orderly development of our knowledge of the retinenes. Synthesis also was prominent in the resolution of the difference of opinion—to the extent that it has been resolved—between the Harvard and Liverpool schools on the mode of attachment of the chromophore to the protein. This controversy arose from problems connected with indicator yellow—one of the products of the bleaching of rhodopsin. Figure 3 shows the synthetic compound retinylidenemethylamine, an indicator yellow analogue formed as the Schiff base of *all-trans* retinene<sub>1</sub> and methylamine. In the form shown its  $\lambda_{\max}$  in ethanol is 365 millimicrons. Upon acidification,  $\lambda_{\max}$  increases to 440 millimicrons. Pitt, Collins, Morton, and Stok (1955) attributed the acid shift to protonation of the nitrogen. In the case of true indicator yellow, their theory would replace the methyl group with protein—ostensibly scotopsin. Both the Harvard and Liverpool groups now appear to accept the Schiff base as the type of bond between retinene and opsin. It must be pointed out, however, that the acid shift found with such compounds is not in itself big enough to account for the rhodopsin shift of 120 millimicrons reported by Hubbard and St. George (1958). Moreover, these Schiff bases do not bleach with light.

A number of theorists have suggested that differences in absorption maxima of various visual pigments may be attributed to different degrees of extension of the conjugation of the chromophore into the protein. The identification of appropriate amines in the protein re-

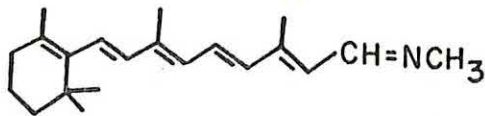


FIG. 3. Retinylidenemethylamine, the Schiff base of retinene and methylamine.

mains a formidable task. Most investigators have assumed, like Dartnall (1957), that the principle components of opsin—with  $\lambda_{\max}$  around 290 millimicrons—are such amino acids as tyrosine and tryptophane. The organic chemist Oster (1958) has stated that the only protein residues identifiable by their ultraviolet spectra are tyrosyl and phenylalanyl, and no additional conjugated structures. He regards extension of conjugation by the protein as therefore unlikely. In the author's laboratory, however, it has been shown that a number of nucleoside phosphates—derivatives of nucleic acids—have sharp absorption peaks in this general vicinity. Moreover, their involvement in the visual process has been indicated by the recent discovery (McConnell & Scarpelli, 1963) that rhodopsin extracts contain substantial amounts of a light- and wavelength-sensitive enzyme which splits off phosphate from nucleoside triphosphates—preferentially adenosine triphosphate. The liberation of each phosphate in the typical reaction of this kind releases up to 11,000 calories of energy per mole of adenosine triphosphate, and may provide the energy for driving other reactions involved in the propagation of the visual sensation.

As a compromise in the controversy over whether the retinene-opsin binding was Schiff base, as claimed by the Liverpool group, or sulfhydryl, as Wald had earlier claimed, Dartnall (1957) proposed that the sulfhydryl groups Wald detected in bleached solutions were exposed by the uncovering of the opsin during isomerization of the retinene. In this way the Schiff base bond is allowed, but the appearance of additional sulfhydryl groups upon bleaching could still be explained. In our own laboratory it has recently been shown that opsin contains sulfhydryl groups which behave, in the presence of known blocking

reagents, very much like those associated with the adenosinetriphosphatase activity of myosin—the muscle protein. It also appears that the activity of opsin in the splitting of adenosine triphosphate is reduced by removal of retinene, or is enhanced by incubation with exogenous retinene. The basic enzyme appears to be opsin, and the function of the retinene may be twofold: as a cofactor for the efficient splitting of adenosine triphosphate, and as a light-trap when the enzyme is illuminated. It is not possible at the present time to prove or disprove the "unmasking" proposal of Dartnall. However, if it is true, the increase in sulfhydryl groups brought on by bleaching would have to reduce the activity of the enzyme, at least up to a certain point, in order to be consistent with our data.

At Michigan State University, Barnett Rosenberg (1962) has reported experiments on semiconductors made of two sheets of conductive glass between which lies a glasslike preparation of  $\beta$ -carotene—the prototype of the chemical family to which retinene belongs. These experiments are of great interest because results so far indicate the possibility that a single chromophore of this chemical family may give a different photoconductive response for every wavelength absorbed. When a voltage is impressed across the sandwich, incident light causes a photocurrent the magnitude, polarity, and distribution in time of which are determined by the wavelength of the incident light. In Figure 4 are plotted on the same time axis the photocurrents produced by four different lights in a  $\beta$ -carotene glass cell across which was applied 300 millivolts. Note that the polarity of the response to yellow and red lights is opposite to that of the blue and green responses. Proper adjustment of intensities of any two lights which separately produce opposite

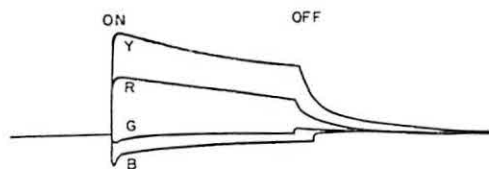


FIG. 4. Photocurrents resulting from application of lights of four different colors to a  $\beta$ -carotene preparation. (Ordinate: current; abscissa: time. Courtesy of B. Rosenberg.)

polarities of response produces a response to both lights, simultaneously presented, which is comparable to the response to a single light of intermediate wavelength. Although the particular relation obtained between wavelength and resulting photocurrent holds only for the restricted geometry of the sandwich he chose, the implications of this research, like that of our enzyme research, for psychophysical theory are large and inescapable. Rosenberg has already tentatively related his data to those of Svætichin (1956) and of MacNichol and Svætichin (1958), from which it appeared that insertion of microelectrodes to certain positions in isolated fish retina permitted recording of a response which reversed polarity as wavelength increased. It must be remarked that MacNichol and Svætichin showed the probable location of the microelectrode tip for this particular kind of response as being in the inner part of the inner nuclear layer, relatively far removed from the photopigment. Such reversals of polarity could be explained by neural hypotheses as well as by chemical ones. And as might be expected, psychologists and electrophysiologists have already begun to make war on Rosenberg's inferences. Nonetheless, his data demand being related to the possible chromophore in the living retina, and Rosenberg approaches this task with refreshing independence.

Some other facets of his theory, while not particularly upsetting to psycholo-

gists, are challenging to previous models for vision. The possibility must be considered that bleaching, including isomerization, does not supply the initial energy for the visual process at all, but plays a nonetheless important role in removing from the range of energetic excitability large chunks of the photopigment. So regarded, bleaching would be superfluous for photopigments with relatively high activation thresholds, and we should therefore not be surprised by the presence in the retina of unbleachable visual pigments, probably in cones. In such systems, whether they bleach or not, the energy would be photoconductive and continuous. If, as Rosenberg suggests, it is also wavelength dependent, an almost indefinitely large variety of responses could be expected from one lowly receptor.

Our enzyme research shows that when opsin is extracted from bleached outer segments of cattle retinas, in which the un-isomerized retinene must be considered at a minimum, the activity of the enzyme is still strong, discriminates light from dark, but lacks the wavelength sensitivity of rhodopsin. This is regained in the presence of added retinene. Electron micrographs show the site of this enzyme to be the outer segment of both rod and cone (Figure 5). If the color sensitivity of a single receptor is adequately represented by the extract we work with—an assumption always in question but never disproved—a single receptor would be capable of color discriminations.

Lastly, let us consider briefly the approach of Rushton (1955, 1958c) who directs light at the retina in the living organism, and measures changes in the density of the bleachable pigments by comparing the reflected energy of the bleaching light with that of a nonbleaching light. This is a difficult measurement, particularly in some species which

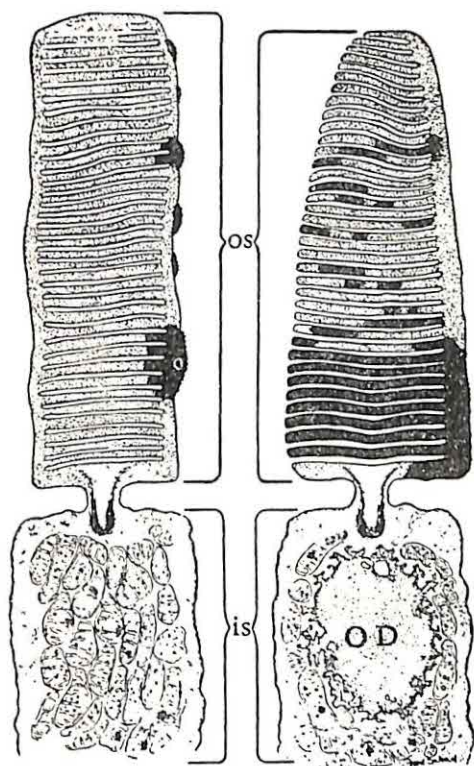


FIG. 5. Artist's representation of electron-microscopic distribution of adenosine triphosphatase in outer segments of red rod (left) and principal cone (right) in the frog retina. (Courtesy of D. G. Scarpelli and E. L. Craig.)

are heavily pigmented, including man. The density of rhodopsin at various points on the dark-adapted retina as determined by this method shows good correspondence with the histological distribution of the retinal rods. Both rhodopsin and rods are, as expected, absent in the fovea and the optic papilla. Like the correspondence reported by Crescitelli and Dartnall between spectral sensitivity of the dark-adapted human eye and the density spectrum of human rhodopsin, these data strongly imply one dark photopigment—rhodopsin.

In the photopic case, Rushton reported identification of two bleachable pigments—chlorolabe and erythrolabe, whose  $\lambda_{\max}$  are in the green and red,

respectively. Chlorolabe, moreover, was the only pigment shown by this technique in the retina of the protanope, since the difference spectrum obtained was independent of the wavelength of the bleaching light. The difference spectrum also agreed well with the protanope's spectral sensitivity. An examination of the fovea of a deuteranope, and another of a deuteranomalous fovea, disclosed only about 10% of the pigment density of the protanope—too little on which to make determinations. Identification of the third pigment generally accepted as required to make color matches has not yet been reported by Rushton.

The method, despite its cleanness, suffers from two obvious limitations. One is psychological: threshold determinations are made by the traditional verbal report of the subject—which it would be difficult to represent to a chemist as indicating absorption or nonabsorption of light. The second is that the method applies only to pigments which will bleach. Whether the blue cone pigment is unbleachable in the human retina, or some other problem prevents its identification by this method, is not at present clear. Suggestions that the blue-sensitive pigment is actually the rod pigment have not met with much favor among most investigators. Our own finding that a photostable pigment with  $\lambda_{\max}$  about 420 millimicrons regenerates from cattle or rabbit opsin and retinene, together with Dartnall's report (1960) of a pigment extracted from the pure-cone retina of the gray squirrel which had an absorption maximum at 502, are only two of several contrasting facts which require that the question be left entirely open.<sup>1</sup>

<sup>1</sup> Since this manuscript was written, Marks (1963) and E. F. MacNichol (personal communication, 1963) have reported absorption spectra for red, green and blue isolated, intact

The differentiation, in the same extract, of scotopic and photopic pigments appears to this author to be at one and the same time the most frustrating and the most promising avenue of research into the chemistry of the receptors. Evidence is now mounting that the system as we find it in extracts is a far cry from that in the receptor. A particularly important example is illustrated by our present inability to extract from the outer segments elements of the cytochrome apparatus which we know from light (McConnell, in press) and electron histochemistry is present to substantial degree in these structures. There is excellent correlative reason to believe the cytochromes are intimately linked with the adenosine triphosphatase enzyme which we can extract. Moreover, it is not too early to suggest that these cytochromes may be associated with the photopic color-sensitivities of the retina. Development of an appropriate extractive technique is of pressing concern.

## REFERENCES

- BOLL, F. Zur Anatomie und Physiologie der Retina. *Monatsber. preuss. Akad. Wiss. Berlin*, 1876, 783-787.
- BROWN, P. K., & WALD, G. Visual pigments in human and monkey retinas. *Nature, London*, 1963, 200, 37-43.
- COPE, F. W., SEVER, R. J., & POLIS, B. D. Reversible free radical generation in the melanin granules of the eye by visible light. *Arch. Biochem. Biophys.*, 1963, 100, 171-177.
- CRESCITELLI, F. The natural history of visual pigments. *Ann. N. Y. Acad. Sci.*, 1958, 74, 230-255.
- CRESCITELLI, F., & DARTNALL, H. J. A. Human visual purple. *Nature, London*, 1953, 172, 195-196.
- DARTNALL, H. J. A. Visual pigment 467: A photosensitive pigment present in tench retinæ. *J. Physiol.*, 1952, 116, 257-289.
- DARTNALL, H. J. A. The visual pigments. New York: Wiley, 1957.
- DARTNALL, H. J. A. Visual pigment from a pure cone retina. *Nature, London*, 1960, 188, 475-479.
- FRIDERICIA, L. S. & HOLM, E. Experimental contribution to the study of the relation between night blindness and malnutrition: Influence of deficiency of fat-soluble A-vitamin in the diet on the visual purple in the eyes of rats. *Amer. J. Physiol.*, 1925, 73, 63-78.
- HAGINS, W. A. The quantum efficiency of bleaching of rhodopsin *in situ*. *J. Physiol.*, 1955, 129, 22P-23P.
- HAGINS, W. A. Flash photolysis of rhodopsin in the retina. *Nature, London*, 1956, 177, 989-990.
- HECHT, S. Photochemistry of visual purple: I. The kinetics of the decomposition of visual purple by light. *J. gen. Physiol.*, 1920, 3, 1-13.
- HECHT, S. Photochemistry of visual purple: II. The effect of temperature on the bleaching of visual purple by light. *J. gen. Physiol.*, 1921, 3, 285-290.
- HECHT, S. Photochemistry of visual purple: III. The relation between intensity of light and the rate of bleaching of visual purple. *J. gen. Physiol.*, 1924, 6, 731-740.
- HUBBARD, R. Retinene isomerase. *J. gen. Physiol.*, 1956, 39, 935-962.
- HUBBARD, R., & KROFF, A. The action of light on rhodopsin. *Proc. Nat. Acad. Sci.*, 1958, 44, 130-139. (a)
- HUBBARD, R., & KROFF, A. The mechanism of bleaching rhodopsin. *Ann. N. Y. Acad. Sci.*, 1958, 74, 266-280. (b)
- HUBBARD, R., & ST. GEORGE, R. C. C. The rhodopsin system of the squid. *J. gen. Physiol.*, 1958, 41, 501-528.
- KÖTTGEN, E., & ABELS DORFF, G. Absorption und Zersetzung des Sehpurpurs bei den Wirbeltieren. *Z. Psychol. Physiol. Sinnesorg.*, 1896, 12, 161-184.
- KÜHNE, W. *On the photochemistry of the retina and on visual purple*. London: Macmillan, 1878.
- McCONNELL, D. G. Enzymes of the retina. In J. H. Prince (Ed.), *The rabbit in eye research*. Springfield, Ill.: Thomas, in press.
- McCONNELL, D. G., & SCARPELLI, D. G. Rhodopsin: An enzyme. *Science*, 1963, 139, 848.
- MACNICHOL, E. J., & SVAETICHIN, G. Electric responses from the isolated retinas of fishes. *Amer. J. Ophthalmol.*, 1958, 46, 26-40.
- MARKS, W. B. Difference spectra of the visual pigments in single goldfish cones. Unpublished doctoral dissertation, Johns Hopkins University, 1963.
- MORTON, R. A., & GOODWIN, T. W. Prepara-
- cones in goldfish. Brown and Wald (1963) have reported the spectra of *in situ* pigments in monkey and human foveas.

- tion of retinene *in vitro*. *Nature, London*, 1944, 153, 405-406.
- MÜLLER, H. Zur Histologie der Netzhaut. *Z. wiss. Zool.*, 1851, 3, 234-237.
- OSTER, G. Models of the visual process. *Ann. N. Y. Acad. Sci.*, 1958, 74, 305-309.
- PITT, G. A. J., COLLINS, F. D., MORTON, R. A., & STOK, P. Studies on rhodopsin: Part 8. Retinylidenemethylamine, an indicator yellow analogue. *Biochem. J.*, 1955, 59, 122-128.
- ROSENBERG, B. Electronic charge transport in carotenoid pigments and a primitive theory of the electroretinogram. *J. Photochem. photobiol.*, 1962, 1, 117-129.
- RUSHTON, W. A. H. Foveal photopigments in normal and colour-blind. *J. Physiol.*, 1955, 129, 41P-42P.
- RUSHTON, W. A. H. Blue light and the regeneration of human rhodopsin *in situ*. *J. gen. Physiol.*, 1958, 41, 419-428. (a)
- RUSHTON, W. A. H. Kinetics of cone pigments measured objectively on the living human fovea. *Ann. N. Y. Acad. Sci.*, 1958, 74, 291-304. (b)
- RUSHTON, W. A. H. Visual pigments in the colour blind. *Nature, London*, 1958, 182, 690-692. (c)
- SVAETICHIN, G. Spectral response curves from single cones. *Acta physiol. Scand., Stockholm*, 1956, 39(Suppl. No. 134), 17-46.
- TANSLEY, K. The regeneration of visual purple: Its relation to dark adaptation and night blindness. *J. Physiol.*, 1931, 71, 442-458.
- TRENDELENBURG, W. Quantitative Untersuchungen über die Bleichung des Sehpurpurs in Monochromatischem Licht. *Z. Psychol. Physiol. Sinnesorg.*, 1904, 37, 1-55.
- WALD, G. Vitamin A in the retina. *Nature, London*, 1933, 132, 316-317.
- WALD, G. Photo-labile pigments of the chicken retina. *Nature, London*, 1937, 140, 545-546.
- WALD, G. The porphyropsin visual system. *J. gen. Physiol.*, 1939, 22, 775-794.
- WALD, G. The biochemistry of vision. *Annu. Rev. Biochem.*, 1953, 22, 497-526.
- WALD, G., BROWN, P. K., & SMITH, P. H. Iodopsin. *Federat. Proc.*, 1952, 11, 304.
- WALD, G., BROWN, P. K., & SMITH, P. H. Cyanopsin: A new cone pigment of vision. *Science*, 1953, 118, 505-508.

(Received March 3, 1963)

## DISCUSSION: COMPETING THEORIES OF RECEPTOR EXCITATION

R. M. BOYNTON  
*University of Rochester*

This discussion is mainly an examination of a theoretical idea suggested by Enoch and McConnell. They suggest that a single photoreceptor may be able to respond so that the output of the receptor is "tagged" according to the wavelength of the quantum absorbed. This idea stands contrary to the classical trichromatic scheme, according to which the probability of quantum absorption depends upon the energy of the incident quantum (inversely related to wavelength), but where once a quantum is absorbed all information pertaining to the wavelength associated with that quantum is forever lost. The trichromatic scheme is reviewed, together with the evidence which supports it. Difficulties raised by some of the suggestions of Enoch and McConnell are then evaluated in this context.

I wish to comment particularly upon a controversial theoretical idea that has been seriously proposed by McConnell, and which also has been mentioned by Enoch. The suggestion is that a single receptor in the eye may be capable of generating a differentiated spectral response, one that is "tagged" in some way so that its response carries with it information about the wavelength of the light absorbed. In other words, it is proposed that a single receptor may "know" what the wavelength of the incident light is. It is also implicit in this argument that the receptor be capable of transferring this differential information about wavelength to higher levels of the visual system.

These speculations stand diametrically opposed to a general principle of color discrimination which may be called the "trichromatic scheme." According to this principle, the responses of receptors are not tagged according to the wavelength of the light that excited them.

I have decided to make it my principal responsibility in this discussion to: (a) describe what the trichromatic scheme is; (b) indicate the kind of experimental evidence that has made it the favorite scheme of the color theorist for more

than a century; and (c) with *a* and *b* as background, point out the implausibility of the tagged-response theory of color discrimination. Before doing so, however, it will be useful first to consider the simpler case of rod vision.

### ROD VISION

When the maximum luminance level of environmental objects is low—below about 5 or 6 microlamberts—there is no color discrimination, and all vision (excluding the effects of memory color) is black and white. The generally-accepted reason for this is that scotopic vision is mediated by only one class of receptor—the rod. Either all rods have the same spectral sensitivity, or, if they do not, there is no separation of the signals transmitted by those having one type of spectral sensitivity and those having another. In either case, the stimulation of rods by light of any wavelength leads to the same sensation of scotopic white which is produced by stimulation by any other wavelength, providing that relative intensities are suitably adjusted. When many quanta are incident upon the eye, the resulting sensation depends simply upon the rate at which quanta are being absorbed by a collection of rods

(for any fixed condition of adaptation). Not only does the rod not know where, along its length, the absorption of a given quantum occurred, but the optic nerve fiber to which a collection of rods is synaptically connected has no information concerning in which particular rods the quanta were absorbed. Moreover, and this is most important for the present argument, the receptor also does not know what the wavelengths of the absorbed quanta were.

Whether or not a given quantum will be absorbed is unpredictable, being a matter of probability. Some quanta, however, stand a much greater probability of absorption than do others, depending upon their energy. For example, suppose that the maximum probability of a quantum being absorbed in some receptor is about .20. This maximum probability will occur at the peak wavelength of scotopic sensitivity, at a wavelength of about 505 millimicrons. At this wavelength, a quantum has an energy value of slightly less than  $4 \times 10^{-12}$  erg (see Table 1). Any quantum possessing more or less energy than this is less likely to be absorbed. For example, a quantum having an energy of  $3 \times 10^{-12}$  erg (corresponding to about 660 millimicrons in the red) will stand less than one chance in 10,000 of being absorbed.

However—and this is the main point—once absorption has taken place, the effect produced by one quantum is the same as that produced by another. The molecule of rhodopsin that absorbed the quantum will have reacted in an all-or-none manner without receiving any information about what the energy of the quantum was.

The psychophysical evidence in support of the point of view just presented lies mainly in the experimental work underlying the relation:

$$\int E_{1\lambda} V'_{\lambda} = \int E_{2\lambda} V'_{\lambda}.$$

Here  $V'_{\lambda}$  is the scotopic luminous efficiency function, and  $E_{1\lambda}$  and  $E_{2\lambda}$  are taken to be any two distributions of radiance in fields which match. For the standard observer, any two fields which satisfy this equation will match, all will have the same scotopic white color, and all potential information about the hue and saturation of objects is lost.

#### THE TRICHROMATIC SCHEME

The main point about the trichromatic scheme that I wish to emphasize is that it implies the same assumptions for cone vision that have just been reviewed for rod vision, except that in this more complex case there must be three classes of photopigment. One class has its maximum absorption in the red, meaning that quanta having an energy of about  $3.3 \times 10^{-12}$  erg are most likely to be absorbed, with the probability falling off for quanta of higher or lower energy. Another class has its maximum probability of absorption at about  $3.7 \times 10^{-12}$  erg in the green, with the third at about  $4.6 \times 10^{-12}$  erg in the blue. A given quantum (say in the green) might be absorbed by any one of the three kinds of photopigment, but is more likely to be absorbed in the "green" kind than in either of the others. Once absorbed, regardless by which kind of pigment, the receptor in which that pigment is located

TABLE 1  
EQUIVALENT WAVELENGTH, ENERGY PER QUANTUM, AND ASSOCIATED COLOR SENSATIONS  
FOR THE NEUTRALLY-ADAPTED EYE

Wave-length	Energy per quantum	Color sensation
400	$4.96 \times 10^{-12}$ erg	Violet
450	$4.41 \times 10^{-12}$ erg	Blue (slightly reddish)
500	$3.97 \times 10^{-12}$ erg	Bluish green
550	$3.61 \times 10^{-12}$ erg	Yellowish green
600	$3.31 \times 10^{-12}$ erg	Orange
650	$3.06 \times 10^{-12}$ erg	Red
700	$2.83 \times 10^{-12}$ erg	Red

does not know: (a) what energy level that quantum possessed, (b) in which particular molecule along the length of its outer segment the quantum happened to be caught, or (c) in which type of pigment molecule the absorption took place (if, as is theoretically possible, there should be a mixture of pigment types within the receptor).

The trichromatic scheme is not concerned with details about the processing of chromatic information beyond the level of photopigments, but perhaps a few comments about the great beyond are nevertheless in order. The trichromatic scheme is much more restricted in its scope than is a complete trichromatic theory of color vision. The simplest trichromatic theory states that there are three cone types (one for each pigment), three classes of optic pathway (one for each receptor), and that color sensations depend somehow upon the blending of these three classes of signal at the brain. One can, however, be a supporter of the trichromatic scheme and nevertheless adopt much more complicated ideas concerning what happens at subsequent stages in the visual system. For example, the evidence for an "opponent colors" transformation is now very strong, but the idea of such recoding of input information does not in any way alter the validity of the concepts involved in the basic trichromatic scheme. I have commented more extensively on this elsewhere (Boynton, 1960).

#### EVIDENCE SUPPORTING THE TRICHROMATIC SCHEME

The one overwhelmingly important reason why the trichromatic scheme has been so popular is that it easily accounts for the remarkable facts associated with color mixture. Briefly stated, color mixture concerns the fact that there are many pairs of physically different stimuli which, when directly compared, match

exactly for hue, brightness, and saturation—that is, for color. Such pairs are called metameric pairs, or metamers. When a systematic effort is made to determine the number of variables that will allow the normal observer to match any color, the number turns out to be exactly three—no more, and no less. Thus any color may be matched by a suitable mixture of exactly three primary colors, with the choice of primaries being arbitrary within these bounds: (a) no mixture of two of the primaries can match the third, and (b) in certain cases, one of the primaries may be added to the side of the colorimetric matching field that contains the color to be matched, being treated as if it were a negative amount of light added to the other side.

Another important series of facts about color vision are known as "Grassman's laws." These can be variously stated, but I think this is the simplest way: there exists an approximate isomorphism between the empirical system of color-matching experiments and the formal system of ordinary algebra. If we replace the "=" sign of algebra by the "matches with" operation of colorimetry, the "+" sign of algebra by the optical superposition of lights in the colorimeter, and if we take the "-" operation to imply, as previously noted, that the stimulus in question is on the "wrong" side of the matching field, then (to the extent that Grassman's laws are really true) the formal and empirical systems are isomorphic. This permits one to calculate many possible experimental results without doing each and every experiment. It also becomes possible to transform from one system of primaries to another. The facts of color mixture have been collated and standardized in an internationally recognized colorimetric system, in which the Commission Internationale de l'Éclairage chromatic-

ity diagram is perhaps the crowning achievement (LeGrand, 1957, Ch. 8). In this rather widely-misunderstood diagram, all stimuli which are metameric with respect to each other plot at a common point, all color mixtures fall along straight lines, and the spectral colors appear along a partially-closed curved locus within which all real mixtures must lie.

The trichromatic scheme accounts, at one fell swoop, for the whole constellation of facts about color mixture that have just been described. It also accounts, if one assumes the pigment concentrations to be low, for the resistance of metameric matches to alteration by adaptation or changes in luminance level. The three-pigment hypothesis of the trichromatic scheme accounts for these facts as follows. If two stimuli match exactly in all respects, then it is supposed that the rate of quantum absorption by each of the three kinds of photopic visual photopigments is exactly the same for both stimuli. Unless something happens to alter the spectral sensitivity exhibited by the receptors containing these pigments, any pair of metameric stimuli behave as they do simply because they exert exactly equivalent effects at this very peripheral level of the visual system. They are indeed "physiologically identical" and from this point on, it matters not what complex transformations may occur in the encoding of the information, how many receptor types there are, or how much of which pigment is in each receptor.

Now both Enoch and McConnell have speculated that the single receptor may not merely instigate an undifferentiated signal correlated with the quantity of absorbed light, but that the job of wavelength discrimination may be handled in part by, and within, the individual receptor. Rosenberg's model, for example, suggests that the sign and temporal char-

acter of the response of a given type of photopigment might depend upon wavelength. The model is completely antithetical to the somewhat updated view of the classical scheme that I have tried to describe here and, it seems to me, raises far more questions than it answers. For example:

1. Exactly how can the trichromacy of human color vision be accounted for by this model?
2. How can the information related to wavelength be retained and transmitted to the higher visual centers, along with information related to intensity?
3. Where do Grassman's laws come from?
4. Why is no information about wavelength transmitted by the scotopic system?

Another body of experimental data which has been to some extent accounted for by the trichromatic scheme, though not without difficulty, is chromatic adaptation (MacAdam, 1961; Stiles, 1961). Time simply does not permit me to go into the matter. Let me say, however, that I cannot see any way that the Rosenberg type of model can even begin to account for the experimental facts.

#### DIFFICULTIES RAISED BY ENOCH'S WORK

I would like now to consider some of what Enoch has said, from the standpoint of the trichromatic scheme. After considering the possibility that the individual receptor might generate a differentiated response depending upon wavelength, he has admitted that the undifferentiated response is much more probable. But there are problems for the trichromatic scheme raised by his important experiments concerning waveguide effects in receptors, from which work it seems quite clear that the transmittance of a given retinal cone is by no

means independent either of the wavelength or the angle of the incident light. Such waveguide effects would have the same consequence, in principle, as the interposition of directionally-sensitive color filters in front of the separate receptors. It is quite correct that cones with three such filters and only one kind of photopigment could conceivably do the job of color discrimination very well and be entirely consistent with the trichromatic scheme. The key question, which Enoch has implied but perhaps has not stated straight out, concerns whether the selective wavelength effects that he observes are sufficiently stable and of the right quantitative sort to allow these to serve as the "filters" of human receptors containing only one photopic photopigment. To me, this surely seems unlikely: first, there is the experiment of Brindley and Rushton, which Enoch mentioned, where monochromatic light was put through the sclera to stimulate the receptors from behind, with no color change reported when compared to ordinary stimulation. One can also observe a similar lack of gross color change when observing monochromatic scattered light in the eye, or that reflected near the optic disc. Finally, the effects of change of angle of incidence and wavelength are very similar, so far as the waveguide modes are concerned, so that it is difficult to understand how the receptor could ever sort them out.

On the other hand, there are a number of instances which I have deliberately side-stepped to this point, where the trichromatic scheme breaks down. One of these concerns the failure of metameric matches to hold at high levels or after intense adaptation. Another concerns the inability of subjects actually to make perfect color matches since such matches are often not perfect, even though the best match that the subject

can make is highly reproducible, as psychophysical judgments go. A third failure of the trichromatic principle concerns the Stiles-Crawford effect of Type 2 (Enoch & Stiles, 1962), which requires the postulation of a change of spectral sensitivity in at least one mechanism depending upon the angle of incidence of light upon the receptor although the change required is not large. A fourth problem concerns the failure of some of Grassman's laws to hold, particularly in peripheral vision. Such effects as these seem to be variations on the main theme. They seem to be saying: "The trichromatic scheme is almost true, but actually it is only an approximation. There must be complications." Enoch's work shows us where some of these complications probably lie.

The facts of color mixture depend upon the stimulation of many thousands of receptors. Color discrimination is very poor in small fields and improves with area at a much faster rate than luminance discrimination (MacAdam, 1943). If there were no correlation between the "filtering" action caused by variations in the dimensions of the tiny receptors and the photopigments contained therein, then psychophysical investigation with reasonably large areas would give us information closely related to photopigment sensitivities, despite the complications introduced by the waveguide effects. The only requirement is that the numbers of receptors be sufficiently large for random effects to cancel out. This suggestion, which is one of the possibilities considered by Enoch, seems most probable to me. It is compatible so far as I can see with other evidence.

One can imagine that Nature said, as she went along her evolutionary way, Let's build a color-discriminating system. But, in order to keep the good spatial resolution that we developed during the past millenium (which, after all, is the capacity of paramount

importance in vision), the receptors must be very small. This in turn interferes with the stable spectral sensitivity of receptors containing a given kind of photopigment. How do we solve this problem? [a] By making color discrimination depend upon the stimulation of many receptors; [b] by dissociating color vision to a large extent from the spatial acuity task; [c] by creating three or more classes of photopic receptors, with each member of a class containing exactly the same pigment or ratio of pigments, and then by causing the initial coding of color information to be in terms of the relative activation of these various receptor types.

## REFERENCES

- BOYNTON, R. M. Theory of color vision. *J. Opt. Soc. Amer.*, 1960, 50, 929-944.
- BRINDLEY, G. S., & RUSHTON, W. A. H. The colour of monochromatic light when passed into the human retina from behind. *J. Physiol.*, 1959, 147, 204-208.
- ENOCH, J. M., & STILES, W. S. The colour change of monochromatic light with retinal angle of incidence. *Opt. Acta*, 1962, 8, 329-358.
- LEGRAND, Y. *Light, colour, and vision*. New York: Wiley, 1957.
- MACADAM, D. L. Color discrimination and the influence of color contrast on visual acuity. *Kodak Res. Lab. Rep.*, 1943, No. 1105.
- MACADAM, D. L. A nonlinear hypothesis for chromatic adaptation. *Vis. Res.*, 1961, 1, 9-41.
- STILES, W. S. Adaptation, chromatic adaptation, colour transformation. *An. Real Soc. Esp. Fis. Quimico*, 1961, 57, 149-175.

(Received March 3, 1963)

## DISCUSSION: COMPETING THEORIES OF RECEPTOR EXCITATION

H. RICHARD BLACKWELL

*Ohio State University*

The opinion is expressed that the classical theory that the color receptors of human vision contain rhodopsin-related photopigments must be abandoned in part, in view of evidence presented by Enoch and McConnell. Additional evidence against the classical theory is presented, derived from luminosity data on patients with retinal abnormalities which resulted in loss of all but 1 color system. The view is expressed that the color receptors are each incapable of differentiated response to equal amounts of absorbed energy of different wavelengths.

What seems to me the most important purpose of the present symposium has been well served by the presentations prepared by Enoch and McConnell. Enoch and McConnell have presented what must be conceded are unconventional theories of receptor excitation in the retina. It may be said that both have presented alternatives to the classical theory that the action spectra of the color receptors may be understood as absorption spectra of rhodopsin-related photopigments. They have at least described complications or difficulties with the classical theory. Enoch has presented evidence that the optical properties of receptors influence their response to different wavelengths, and McConnell has presented evidence that the Wald model of excitation by breakdown of rhodopsin is inadequate. Both have perhaps gone unnecessarily far in suggesting that a single receptor may be capable of giving differentiated responses to different absorbed wavelengths other than those due to differences in the absorbed quantity of energy. Boynton has designated the notion that a single receptor cannot respond differentially to wavelength as the trichromatic scheme and has summarized well the convincing evidence in support of this assumption.

Let me present one additional line of

evidence which, I believe, limits the range of possible theories of retinal receptor excitation. Blackwell and Blackwell (1959, 1961) have shown that patients with diseases and developmental abnormalities of the retina reveal the fundamental action spectra of the three classes of receptors required for the trichromatic scheme, since aberrant retinas seem to exhibit simplified color systems consistent with the view that one or more of the color receptors has been rendered inoperative. Most dramatic are the patients classified as blue monochrome monochromats, who have spectral responsivity only in the region usually identified with the blue-sensitive color receptor.

Assuming as we do that in this case we may equate the luminosity curve to the action spectrum of the blue color receptor, we may test the hypothesis that the cone photopigment is rhodopsin related. Figure 1 shows the expected rhodopsin-related photopigment derived by the method of Dartnall (1957) with the action spectrum from the patients' luminosity curve, corrected for spectral absorption of the ocular media and recomputed in terms of an equal quantal spectrum, presented for comparison. The action spectrum is clearly more narrow than is to be expected on the basis of

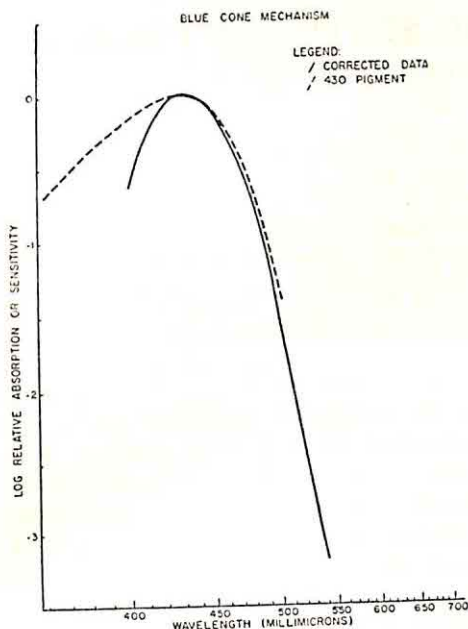


FIG. 1. Comparison of inferred photopigment absorption for the blue-sensitive cone mechanism (solid curve) with the computed absorption curve for the rhodopsin-related photopigment with a 430 millimicrons maximum (dashed curve).

a rhodopsin-related pigment. We have obtained strong suggestions that the action spectra of red- and green-sensitive receptors are also more narrow than is to be expected for rhodopsin-related photopigments.

These results can mean either that

Enoch's waveguides produce a narrowing of spectral cutoff for color receptors or that McConnell is correct that the spectral responsivity of the color receptors is not at all determined by rhodopsin-related photopigments, or both. Yet, these data provide no difficulty with regard to the trichromatic scheme. My summary of all that has been said is simply that the trichromatic scheme is probably essentially correct, but that we cannot as yet identify either the precise action spectra of the color receptors nor their physical or chemical determinants. However, Enoch and McConnell do present convincing evidence that the classical view that the color receptors contain rhodopsin-related photopigments whose action spectra determine color response cannot be maintained without modification.

#### REFERENCES

- BLACKWELL, H. R., & BLACKWELL, O. M. Identification of rod and cone receptor mechanisms in atypical congenital achromatopsia. *Arch. Soc. Amer. Ophthalm. Optom., Bogota*, 1959, 2, 73-94.
- BLACKWELL, H. R., & BLACKWELL, O. M. Rod and cone receptor mechanisms in typical and atypical congenital achromatopsia. *Vis. Res.*, 1961, 1, 62-107.
- DARTNALL, H. J. A. *The visual pigments*. Wiley: New York, 1957.

(Received March 3, 1963)

## NOTE ON THE SCIENTIFIC STATUS OF FACTORS

JOHN E. OVERALL  
Kansas State University

The frequently assumed relationship between factor analysis results and primary dimensions of objects is seriously questioned. Demonstration analyses are contrived to show that simple structure factors may be quite complex with regard to the conceptually primary dimensions of objects from which measurements were obtained and to indicate that there is no necessary correspondence between the number of factors obtained and the number of primary dimensions of the objects. Recommendations are made for a more realistic view of this valuable methodologic tool.

Over the last decade and a half, a number of methodologists have argued the cause of factor analysis as a method for identifying and understanding the structure of nature. Beginning with Thurstone's (1947) box problem a series of demonstration analyses have been undertaken with the objective of showing correspondence between factors and "real characteristics" or "primary dimensions" of the things being measured (Barlow & Burt, 1954; Cattell & Sullivan, 1962). In a recent issue of the *Bulletin*, Cattell and Dickman (1962) once more have attempted to demonstrate a relationship between factor results and the "incontrovertably known factor structure" of the measurement domain. In view of the persuasive nature of these demonstration analyses, it seems important that a dissenting voice be heard.

Factor analysis has been and will continue to be one of the most fruitful methodologic approaches available for psychological research. It provides a technique for reducing a large multivariate set of correlated measurements to a smaller number of perhaps more meaningful uncorrelated variates. The results need not and do not have inherent in them any necessary relationship to "real" or "primary" characteristics of the objects or persons being measured.

Demonstration analyses are easily contrived for almost any purpose. Previous analyses have been loaded to yield results corresponding to preconceived primary dimensions. It is just as easy to load examples in such a way that results do not correspond to what we conceive to be primary characteristics of objects. Logically, the demonstration analysis has its proper place as a counter example. A single exception will refute the general argument, but no small number of demonstration analyses can prove its truth. The important thing to consider is that if factor analysis results sometimes fail to correspond to primary dimensions of the objects being measured such a failure may occur when we attempt to use it as a method for discovering the "real structure of nature" in areas where that structure is unknown.

### *Simple Structure and Complex Factor Dimensions*

Demonstration analyses of boxes, bottles, geometric figures, cups of coffee, and balls have unerringly resulted in factors whose number and nature correspond to preconceived primary dimensions of the objects being measured. From all of this, what else can the student conclude but that factor analysis is a methodology capable of results which are somehow magically and fundamentally representative of "reality?"

The inference is certainly that one can take any set of complex measurements and come up with simple structure factors which correspond to preconceived primary dimensions. This is just not so. Rotation to simple structure can be understood as an elaborate approach to cluster analysis. It identifies clusters of tests which measure the same things, but there is no assurance that these "same things" are simple and primary dimensions.

Consider a "book problem." Measurements of height, width, and thickness were obtained on each of 100 books. Equations defining 12 complex measures as linear functions of the conceptually primary dimensions of height, width, and thickness are presented in Table 1. It will be noted that the first nine measures are complex combinations of the three "primary" physical dimensions and that the last three are pure measures. Intercorrelations among the 12 measures are presented in Table 2.

A principal axes factor analysis yielded three factors. Orthogonal rotation procedures were employed in an attempt to obtain a simple structure solution disregarding the three pure marker variables which were included in the analysis only to aid in interpretation. The

loadings of the marker variables on the rotated factor axes were then calculated. Loadings of all 12 variables on the three unrotated and rotated orthogonal factors are presented in Table 3, and the transformation matrix used in obtaining the rotated loadings is presented in Table 4. Those who favor oblique rotations will note that, while book dimensions of height and width are substantially correlated, the complex measurements in this example were selected in such a way that orthogonal axes provide adequate representation of the test vector configuration.

Examination of the projections of the pure marker variables (10, 11, and 12) on the rotated orthogonal axes leads to the conclusion that, while the pattern of factor loadings reasonably approaches simple structure for the complex variables, these simple structure factors are highly complex with regard to primary dimensions of height (Variable 10), width (Variable 11), and thickness (Variable 12). Rotated Factor A correlates positively with all three conceptually primary dimensions. Variables 1, 2, and 3 have high and rather pure loadings on Factor A. Examination of the structure of these complex variables as shown in Table 1 indicates that variables with high loadings on Factor A are complex linear combinations of all three conceptually primary dimensions.

Examination of the loadings of the pure marker variables on rotated Factor B indicates that height (Variable 10) and width (Variable 11) have approximately equal negative projections while thickness (Variable 12) has high positive projection on Factor B. Variables 4, 5, and 6 have high loadings on this factor. From Table 1 it can be observed that these variables represent weighted contrasts between height and width, on the one hand, and thickness on the other. Once again, the factor

TABLE 1  
WEIGHTS DEFINING 12 COMPLEX BOOK  
VARIABLES

$X_1 =$	$1.0H + 1.0W + 2.0T$
$X_2 =$	$1.5H + 1.0W + 1.5T$
$X_3 =$	$2.0H + 2.0W + 2.0T$
$X_4 =$	$-0.7H - 0.2W + 3.0T$
$X_5 =$	$-0.8H - 1.2W + 4.0T$
$X_6 =$	$-1.7H - 1.0W + 5.5T$
$X_7 =$	$1.8H - 1.3W - 0.3T$
$X_8 =$	$1.5H - 1.3W - 0.3T$
$X_9 =$	$1.5H - 1.5W + 0.0T$
$X_{10} =$	$1.0H + 0.0W + 0.0T$
$X_{11} =$	$0.0H + 1.0W + 0.0T$
$X_{12} =$	$0.0H + 0.0W + 1.0T$

Note.—H = height; W = width; T = thickness.

TABLE 2

INTERCORRELATIONS AMONG 12 COMPLEX BOOK VARIABLES

	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00	.98	.97	.23	-.03	-.01	.14	-.11	-.16	.82	.80	.61
2		1.00	.99	.05	-.21	-.19	.22	-.05	-.15	.91	.86	.46
3			1.00	-.01	-.27	-.25	.18	-.10	-.22	.91	.90	.40
4				1.00	.96	.97	-.22	-.15	.12	-.29	-.32	.91
5					1.00	.99	-.14	-.01	.28	-.48	-.57	.77
6						1.00	-.23	-.10	.19	-.49	-.53	.79
7							1.00	.96	.87	.51	-.07	-.06
8								1.00	.96	.25	-.34	-.11
9									1.00	.07	-.52	.09
10										1.00	.82	.13
11											1.00	.06
12												1.00

appears to represent a discrete set of variables which measure the same things in approximately the same way, but it is definitely complex with regard to the three conceptually primary book dimensions.

Factor C apparently represents a weighted contrast of height and width. Examination of the pattern of factor loadings indicates that the first 6 variables have near zero loadings on Factor C, while variables with high loadings represent weighted contrasts between height and width dimensions. Thus, Factor C also appears to be complex with

regard to the conceptually primary book dimensions.

What has objective factor analysis and rotation to an approximate simple structure achieved? It has identified groups of variables which correlate highly among themselves and which are relatively independent of other variables. The results depend entirely upon the pattern of intercorrelations among the variables, and these intercorrelations in turn depend, not upon variables which correlate highly measuring the same unitary primary dimensions, but upon relative similarities in their representations

TABLE 3

LOADINGS OF 12 BOOK MEASURES ON THREE ORTHOGONAL FACTORS

	Unrotated			$h^2$	Rotated		
	I <sub>0</sub>	II <sub>0</sub>	III <sub>0</sub>		A	B	C
1	.827	.559	.060	1.000	.988	.101	-.065
2	.913	.397	.098	1.000	.993	-.079	-.012
3	.939	.342	.035	1.000	.983	-.145	-.070
4	-.357	.934	.031	1.000	.120	.995	-.062
5	-.587	.798	.133	.999	-.133	.994	.071
6	-.568	.822	.044	1.000	-.115	.998	-.021
7	.219	-.179	.959	1.000	.222	-.187	.958
8	-.062	-.215	.975	1.000	-.039	-.084	.996
9	-.264	-.002	.965	1.000	-.120	.200	.973
10	.944	.042	.328	1.000	.884	-.389	.258
11	.959	.037	-.280	.999	.824	-.448	-.345
12	.057	.992	.112	1.00	.519	.854	-.016

TABLE 4

TRANSFORMATION MATRIX FOR ORTHOGONAL  
ROTATION OF THREE BOOK FACTORS

.876	-.478	-.066
.459	.880	-.124
.117	.078	.990

of the underlying primary dimensions. Variables may correlate highly which tend to reflect in about the same degree several different primary dimensions of the things being measured. In order for the simple structure criterion to result in factors which represent primary characteristics of the things being measured, it is necessary that only a minority of the variables which reflect one primary characteristic to a high degree also reflect any other one primary characteristic to a high degree. When we do not know beforehand what the primary dimensions are, it is difficult to justify the belief that factor analysis will somehow magically point them out to us.

#### *Number of Factors and Primary Dimensions of Nature*

It may be that we cannot count on factor analysis to result in factors which correspond directly to the "real" dimensions of the thing being measured. Are we still justified in assuming that there is some sort of fundamental relationship between the number of factors and the number of primary conceptual dimensions (whatever those are) of the things being measured? This problem of the number of factors is naturally very important to those who view factor analysis as methodology fundamentally related to the structure of nature.

Cattell and Dickman use a theorem derived by Guttman (1954) which says that the lower bound for the number of factors is equal to the number of latent roots of size 1.00 or larger. In the foregoing example, three and only three roots exceeded 1.00. Thus, in at least one

book problem we have come up with the expected number of factors.

Consider Book Problem Number 2. Twelve complex measures defined as functions of the conceptually primary dimensions of height, width, and thickness are presented in Table 5. These 12 complex measurements were obtained for each of 100 books. Intercorrelations among the 12 complex measures are presented in Table 6.

A principal axes analysis was undertaken with unity in the principal diagonal. Four factors with corresponding characteristic roots greater than unity were extracted. The four factors accounted for 98% of total variance. The roots associated with the four factors were 5.36, 4.29, 1.09, and 1.02, respectively. The loadings of the 12 measures on the four unrotated principal axes are presented in Table 7. Since interest here is in the number of factors, no attempt has been made to rotate the axes. The point is that in two different analyses of complex measurements of the same physical objects different numbers of factors were found. Thus, it appears that neither the number nor the nature of primary factors has inherent relationship to the true conceptual dimensions of the objects being measured.

It should be recognized that the

TABLE 5

TWELVE COMPLEX BOOK VARIABLES AS FUNCTIONS OF HEIGHT, WIDTH, AND THICKNESS

$X_1 = H/H \cdot W \cdot T$
$X_2 = H - 0.2W - 0.2T$
$X_3 = H / (0.2W + 0.2T)$
$X_4 = W/H \cdot W \cdot T$
$X_5 = W - 0.2H - 0.2T$
$X_6 = W / (0.2H + 0.2T)$
$X_7 = T/H \cdot W \cdot T$
$X_8 = T - 0.05H$
$X_9 = T / 0.05H$
$X_{10} = H \cdot W \cdot T$
$X_{11} = H + W + T$
$X_{12} = (H^2 + W^2 + 4.0T^2) \} \exp(\frac{1}{2})$

TABLE 6  
INTERCORRELATIONS AMONG 12 COMPLEX BOOK VARIABLES

	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00	.06	.50	.99	.05	.28	.02	-.59	-.60	-.45	-.21	-.14
2		1.00	.02	.09	.62	.26	-.87	-.09	-.25	.29	.80	.82
3			1.00	.44	-.46	-.24	.32	-.69	-.68	-.68	-.57	-.51
4				1.00	.13	.35	-.03	-.58	-.60	-.43	-.15	-.08
5					1.00	.86	-.84	-.15	-.25	.24	.81	.82
6						1.00	-.53	-.48	-.53	-.17	.39	.42
7							1.00	.00	.15	-.38	-.90	-.91
8								1.00	.98	.86	.33	.27
9									1.00	.75	.19	.12
10										1.00	.68	.65
11											1.00	.99
12												1.00

emergence of more than three factors in this example depends upon the inclusion of complex variables which are nonlinear combinations of height, width, and thickness. The factor model assumes that each complex measure is a simple linear combination of primary factor variates, but we have no control over nor insight into the nature of complexities involved in most practical problems. Even Thurstone (1947) included some nonlinear combinations in his famous box problem. In domains where we do not artificially construct our complex variables, nonlinear complexities are

almost certain to occur and these will influence the number and nature of factors obtained. For example, volume may be a meaningful complex book dimension with three additional independent dimensions of height, width, and thickness also present. Cattell and Dickman (1962) include variables such as weight. For equal density objects, weight is directly proportional to volume which in turn is a nonlinear combination of height, width, and thickness dimensions.

#### *Further Comments concerning the Number of Factors*

Humphreys (1962) has described in precise terms the problem of multiple levels of factor results. He suggests that it is useful to conceive of at least three levels of factor generality, all of which can be represented under an appropriate orthogonal transformation. His discussion appears to bring to formal factor theory the concepts of hierarchical analysis presented by McQuitty (1960a, 1960b) in somewhat different context. Guttman (1958) has also departed from the traditional school in suggesting that it is most reasonable to conceive of test scores as being determined by a large number of sources of variance.

The basic factor model represents the

TABLE 7  
PRINCIPAL AXIS FACTORS FROM 12 COMPLEX VARIABLES

	I	II	III	IV	$h^2$
1	-.39	-.70	-.18	.58	1.00
2	.66	-.50	-.51	-.21	.99
3	-.72	-.39	-.49	-.24	.97
4	-.33	-.72	-.13	.59	1.00
5	.75	-.58	.31	-.01	.99
6	.33	-.72	.59	.02	.99
7	-.82	.48	.14	.11	.94
8	.49	.82	-.10	.27	1.00
9	.36	.88	-.01	.27	.99
10	.77	.48	-.19	.26	.93
11	.97	-.20	-.11	.02	1.00
12	.95	-.26	-.14	.04	.99



analysis is a procedure for obtaining from a large set of correlated measurements a relatively small set of uncorrelated, psychologically meaningful linear components which account for most of the variance in the larger set. The three most usual objectives are parsimony, orthogonality, and meaningfulness. Depending upon the relative emphasis placed upon these different objectives, different methods of analysis and rotation should be considered.

Simple structure is also a valuable concept in factor methodology. The arbitrary reference frame may be rotated for the purpose of displaying the major characteristics of the test configuration. Simple structure factors identify clusters of tests which correlate highly and which are relatively independent of other tests. Frequently we can, through examination of the tests, associate with the factor a meaningful pre-existing concept. Sometimes a new concept may need to be formed to represent the association of items. In all of this, there is no need to assume that simple structure factors will correspond to any particular set of fundamental dimensions of the objects nor, for that matter, to assume that such fundamental dimensions exist in nature.

Considering the results of the first book analysis, we see that meaningful concepts can be associated with the simple structure factors in spite of the fact that these concepts do not correspond to what we might consider to be primary physical dimensions of a book. Factor A is a composite of all three physical dimensions. It corresponds to our psychological conception of the size of the book. An increase in any dimension increases the size of the book. Factor B represents thickness

relative to page size. We might formulate a concept of "obesity" in books; that is, thickness relative to frame. Factor C represents "departure from squareness" in cover dimensions. While we may not be as familiar with these concepts and may not be in the habit of perceiving books in terms of size, obesity, and squareness, these independent dimensions appear quite meaningful. In spite of the meaningful factors obtained, the writer would not encourage one to argue that we should stop considering books in terms of height, width, and thickness because factor analysis has shown us that size, obesity, and squareness are the real primary dimensions. Alternative sets of concepts may prove quite useful.

#### REFERENCES

- BARLOW, J. A., & BURT, C. L. The identification of factors from different experiments. *Brit. J. statist. Psychol.*, 1954, 7, 52-53.
- CATTELL, R. B., & DICKMAN, K. A. A dynamic model of physical influence demonstrating the necessity of oblique simple structure. *Psychol. Bull.*, 1962, 59, 389-400.
- CATTELL, R. B., & SULLIVAN, W. The scientific nature of factors: A demonstration through coffee cups. *Behav. Sci.*, 1962, 7, 184-193.
- GUTTMAN, L. Some necessary conditions for common-factor analyses. *Psychometrika*, 1954, 19, 149-161.
- GUTTMAN, L. L. To what extent can communalities reduce rank? *Psychometrika*, 1958, 23, 297-308.
- HUMPHREYS, L. G. The organization of human abilities. *Amer. Psychologist*, 1962, 17, 475-483.
- MCQUITT, L. L. Hierarchical linkage analysis for the isolation of types. *Educ. psychol. Measmt.*, 1960, 20, 55-67. (a)
- MCQUITT, L. L. Hierarchical syndrome analysis. *Educ. psychol. Measmt.*, 1960, 20, 293-304. (b)
- THURSTONE, L. L. *Multiple-factor analysis*. (2nd ed.) Chicago: Univer. Chicago Press, 1947.

(Received October 29, 1962)

# PSYCHOPHYSIOLOGICAL INTERRELATIONS IN THE SOCIAL BEHAVIOR OF CHICKENS<sup>1</sup>

A. M. GUHL

*Kansas State University*

Chickens establish a social order based on domination and subordination by pecking. The interindividual habits are relatively fixed as social inertia and channel various activities in an order of precedence. Increased aggressiveness produced by treatment with an androgen is suppressed by these habits but becomes evident when an injected bird is matched with a total stranger or under a highly competitive situation. Successful revolts within the peck order are related to the degree of extinction of domination or toleration between individuals. Chicks treated with gonadal hormones show similar suppression of altered aggressiveness by social inertia. The passive dominance of males over females facilitates sexual behavior. Low ranking males have their sexual activity inhibited, whereas low rank among hens augments sexual receptivity. Physiological adaptation to social stress at low levels in the dominance order is brought about by the stimulation of the pituitary-adrenal cortex axis as shown by the hypertrophy of the adrenals.

In the course of over 20 years of research in the social behavior of chickens there have been recurring evidences that the habits of domination and subordination have influenced activities regardless of the physiological conditions of the individual birds. These behavioral phenomena have not received the attention that they appear to merit. Beach (1958) has made the comment that psychologists and endocrinologists have worked independently, with the result that the interrelations of their disciplines have remained a matter of speculation. It is the aim of this review to bring some evidence from these two fields for further consideration.

The writer, a zoologist, has viewed behavior as the actions and reactions of the whole organism and as physiology outwardly expressed. Although conditioning, learning, and memory have a physiological basis, they are not well understood as physiological phenomena and therefore the term psychological has a useful connotation within the context of this review.

The observations and experiments cited herein have considered the following factors, among others, which may alter behavior: physical factors such as extremes of temperature; changes in the thresholds of response caused by fatigue and drive reduction following repeated elicitation, and from stimulation that results in what ethologists call "overflow activities"; conditioning and learning; maturation and aging, which include experience; seasonal factors, such as internal rhythms and endocrine effects due to changing day lengths; genetic basis for strain differences and individuality; and dietary deficiencies or avitaminosis.

Social behavior has been defined variously although with somewhat similar denotations (Guhl, 1962b). According to Carpenter (1942) "social behavior refers to the reciprocal interactions of two or more animals and the resulting modification of the individual's action systems [p. 248]." Others prefer certain limitations, such as occur within the same species. The type of social organization within a given species is determined by the behavior patterns characteristic of that species. Since the

<sup>1</sup> Contribution No. 321, Department of Zoology, Kansas Agricultural Experiment Station, Kansas State University, Manhattan.

phenomena discussed here pertain to the common domestic fowl a brief description of the social hierarchy, or peck order, is required.

#### THE SOCIAL ORGANIZATION

Agonistic behavior, which includes attack and escape reactions, forms the basis for the social order in chickens (Guhl, 1962a). When individuals are marked for identification, one can record all social encounters involving the one pecking and the one pecked. When these observations are tabulated it becomes apparent that such domination is unidirectional; that is, peck rights are established, and that all the birds may be ranked according to the number of hens each pecks without any retaliation. This ranking has been called a dominance order, or peck order. In very small flocks the order may be a straight line hierarchy; however, pecking triangles are common and give the order of precedence a geometrical structure.

Repeated pecking reinforces the dominance relations between all pair combination in the flock. Thus special habits are established between the individuals. In time symbolic threats, or the mere raising or lowering of the head may suffice for the maintenance of pair-reaction patterns. This reduction in frequency of pecking or threatening in time is shown in Figure 1 (intensity of agonistic behavior is not illustrated, Siegel & Hurst, 1962). There are two points of particular interest: one, that the dominance order channels the flow of activity by way of precedence in competitive situations and thus precludes fighting and conserves energy; the other concerns the gradient of habits (or attitudes) from that of a high level of domination to a characteristic response of subordination at lowest levels. Social mobility (changes in rank) is unusual and the relative fixity of response pat-

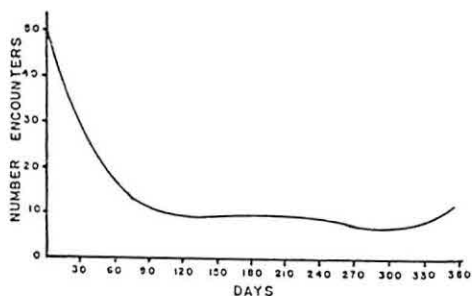


FIG. 1. The decrease in domination after assembly of flocks of hens as shown by total social encounters as a function of time (data from Siegel & Hurst, 1962).

terns results in social inertia within the flock. The mutual interindividual adjustments promote toleration which in turn facilitates the activities within the group (Guhl & Allee, 1944). It is the relation of social habits to the physiology of behavior which will be discussed.

#### GONADAL HORMONES AND SOCIAL INERTIA

*Androgenic Treatment.* Since the early studies of the endocrines suggested a relationship between androgen and aggressiveness, Allee, Collias, and Lutherman (1939) injected testosterone propionate into low ranking hens within several small flocks to investigate its effects. Several of the low ranking, treated birds staged successful revolts and rose in rank, even to top level. However, treated hens in one flock failed to show any mobility in the social order, for which no explanation could be given. Nevertheless, the significance of social inertia was recognized by the statement that "memory and habit reinforce, and may entirely replace the aggressive behavior which is so important in the origin of a social order [p. 436]." During the postinjection period the hens that improved their rank tended to retain their new social status. This was attributed to psychological factors which retarded the rise

but now helped to maintain the newly acquired rank. Similar results were obtained by Allee and Foreman (1955).

During the course of this experiment the treated birds were tested for aggressiveness in initial pair encounters. This test is similar to that used by Maslow (1934a, 1934b). Briefly, when two unacquainted hens are placed into a neutral area they readily establish dominance relations, and relative aggressiveness is indicated by the number of contests won. The treated hens were matched in pairs with normal controls and won most of these encounters, thus indicating that androgen may enhance aggressiveness in the absence of social inertia.

Subsequently some of these birds were used in an experiment on social discrimination. One hen which received a pellet of testosterone propionate showed no social mobility, although she threatened two males when tested in a discrimination cage (Guhl, 1942, p. 145). When she was placed into a pen of strange hens she drove them about the pen (the opposite of what was expected), but on return to her own flock she resumed the behavior associated with her rank (Guhl, Collias, & Allee, 1945, p. 380).

With a flock consisting of 16 individuals, 4 each of 4 breeds, Williams and McGibbon (1956) implanted a pellet of testosterone into 1 of each breed and at several levels in the social order. Observations over a 6-week period failed to show any social mobility. It was concluded that the hormonal effects were too gradual to overcome the already established behavior patterns (social inertia).

These results all suggest that recognition, memory, and habits underlie the social inertia that is responsible for a stable social order. Ill or injured birds may maintain their status provided that

factors for recognition (e.g., combs) do not change abruptly. Successful revolts may occur among normal hens, but rarely in well-managed small flocks. To resolve these apparent discrepancies a study of the conditions under which revolts may be successful or unsuccessful was indicated.

Recent experiments (Guhl, unpublished) have used several flocks of 10 females. Peck orders were allowed to develop to stability and testosterone propionate was then injected daily into certain individuals. The hens to be treated were selected according to the frequency at which they pecked or were pecked. Competitive interactions were stimulated during timed observation periods by presentation of scratch grain and wet mash. Relative social stress was measured in rates of pecking and threatening, which could be used to determine differences in social tension between flocks as well as between individuals within flocks. In competitive situations the injected bird would readily assume a threatening stance, with varying intensity, which in turn would evoke an attack by its ranking penmates. If the social tension was high, then the revolt was unsuccessful, whereas there was a greater chance of success when toleration (tendency toward extinction) was noteworthy. For example, in one set of experiments involving four flocks during several months of observations, there were 292 observed unsuccessful revolts of which 215 were initiated by treated hens. Of only 20 successful revolts, 16 were by treated birds.

It may be concluded that social inertia tends to stabilize the social order irrespective of changes in relative aggressiveness produced by treatment with androgen. Revolts may be successful if the subordinate individual is stimulated sufficiently in a competitive situation

and the existing social relations are marked by toleration.

*Estrogenic Treatment.* The effects of an estrogen on social mobility have received little attention. Allee and Collias (1940) injected estradiol into several hens in five different flocks. Only two shifts in dominance relations were obtained. This was in contrast with the higher upward mobility which resulted when treated with androgen, and contrariwise the status displacement was downward with estrogenic treatment. Williams and McGibbon (1956) failed to obtain any social mobility by the implantation of pellets of diethylstilbestrol. The analysis of recent experiments with estrogen (Guhl, unpublished) have not been complete. No social mobility was recorded. Treated hens showed little change in frequency of pecking although some treated birds appeared to be less intense in their domination. This modification of behavior promoted toleration of subordinate penmates rather than the evocative threatening stance shown after injection with an androgen.

#### DEVELOPMENT OF SOCIAL ORGANIZATION IN CHICKS

In the course of an experiment on selective breeding for levels of aggressiveness (Guhl, Craig, & Mueller, 1960), it was deemed essential to examine the development of aggressiveness and the resultant peck order in chicks (Guhl, 1958). The age, in weeks, at which each chick established definite dominance relations was determined. Normal males began to set up a peck order at 5-7 weeks of age and completed the dominance order at about 8-10 weeks. Females initiated their social positions about 1 week later and required more time to complete the organization. Chicks kept in isolation from hatching onward were assembled when their penned controls formed a

peck order and were able to establish dominance relations in a matter of hours. These results immediately raised the question of the relative effects of maturation and/or learning.

Unisexual groups of males and females were set up as controls and experimental groups from the same hatch with all the chicks in the treated group receiving a gonadal hormone from the second or third day after hatching. Groups of each sex which received an androgen formed their social dominance earlier and in only slightly less time (about 1 week) than their respective controls. Males castrated (capons) at 10 days of age were slow in eliciting aggressive behavior and formed an order 4 weeks later than their controls. But capons treated with an androgen established pair reactions and their peck order earlier than the normal controls. Contrary to expectancy, treatment with an estrogen resulted in both male and female groups showing a similar advanced development of social organization. However, there was a marked difference in response behavior. Estrogen-treated chicks showed little pecking but avoided readily, with the consequence that the social system was essentially a consistent avoidance order. It was somewhat surprising that the hormonal treatments did not induce greater precocity. Could the lag have been due to some form of social inertia, even among immature birds?

To determine whether androgen treatment could induce an earlier formation of peck rights in the absence of learned social responses, five sets of males from the same hatch were used (Guhl, 1958). Two groups were penned from hatching, with one serving as a control and the other as an experimental control receiving androgen. Three sets were placed in spatial isolation and also treated with androgen. The isolates

were assembled at 31, 41, and 51 days of age, and observed for the following 7 days to determine dominance relations.

As may be seen from Figure 2, the experimental controls developed peck rights earlier than the normal controls as expected, and formed 53 out of a possible 55 dominance relations by the end of the observations (56 days), whereas the uninjected chicks established only 36. Treated isolates assembled at 31 days of age began to fight immediately and nearly completed their peck order (42 out of 45) by the end of a week, or at about the time that pen-reared normals began to organize. The other two groups of isolates introduced sexual behavior as a complication. Androgen induces sexual behavior as well as aggressiveness and in the absence of penmates no conditioning could be made in relation to either type of

behavior. Those assembled at 41 days displayed predominately sexual behavior and established only 20 out of 45 peck rights. Chicks assembled at 51 days apparently made some adjustment to the conflicting tendencies for either attack or for sexual approaches. None of the treated birds in flocks showed similar sexual behavior. These results may offer some evidence for the need for spacing, in time, the emergence of behavior patterns during development to permit social adjustments or learning. Treatment with androgen tended to reduce this interval and resulted in a conflict of sexual and dominance drives.

In another experiment of this series, an attempt was made to determine how early in the life of normal chicks that agonistic behavior might occur if there were no social inertia. Four sets of 10 male chicks each were used, with one set in isolation and the other three in pens. Every other day one chick from each of two groups was interchanged. Similarly, a chick from isolation was placed into the third penned group as one was removed and placed into isolation. These two patterns of rotation in the membership of the groups were continued for 8 weeks from hatching. The data in Figure 3 are from daily observations of 15 minutes per pen. All forms of agonistic behavior were not only at a higher frequency in the pen containing chicks from isolation, but also started at an earlier age. Chicks introduced from isolation were a stronger stimulus for attack than those rotated between groups, and differences in their postures were apparent. The results showed that social adjustments may be made early in the life of chicks, even without exogenous androgen.

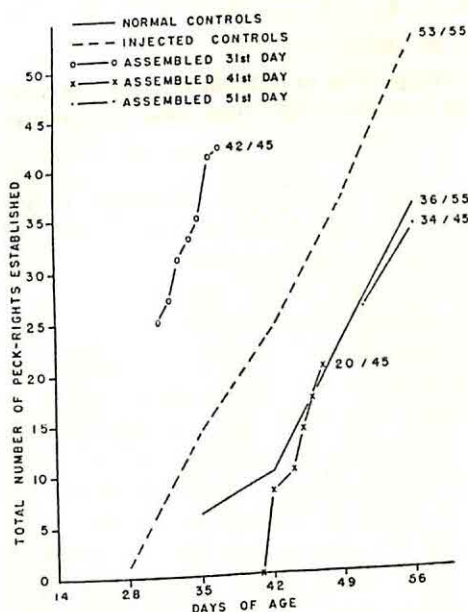


FIG. 2. The total number of peck rights established at various ages among males reared as a group and others reared in spatial isolation and assembled at different ages. (The isolated chicks and one group reared together were injected with androgen. See Guhl, 1958.)

#### SOCIAL STATUS AND SEXUAL BEHAVIOR

A number of early studies on the reproductive behavior among chickens

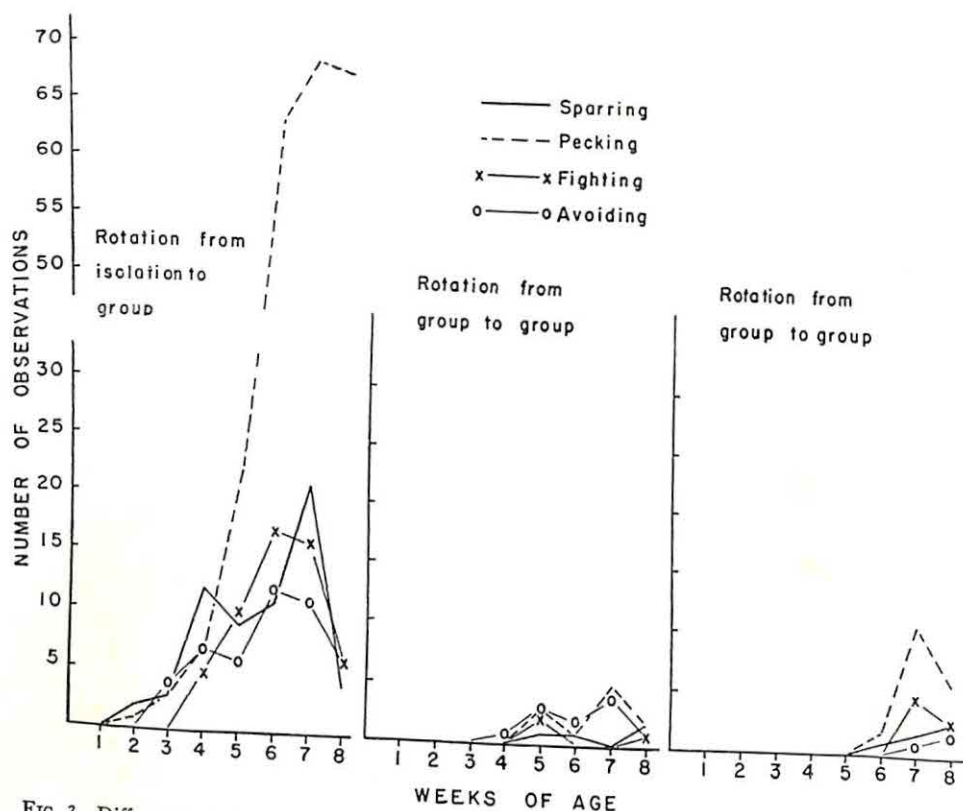


FIG. 3. Differences in the frequency of some behavior patterns of male chicks rotated between isolation and a group and those rotated from group to group. (The latter show social inertia. See Guhl, 1958.)

have shown marked differences between individual males and females in the number of matings which occurred. Some of these suggested a relation to social rank as one possible cause for variability. Since males do not peck the hens, and therefore have a peck order apart from that of the females, the sexes will be considered separately.

**Social Status of Males.** To determine the effects of social rank among males on their sexual activity they were tested singly, and in rotation, with a flock of hens to obtain an estimate of their sexual drive (Guhl, Collias, & Allee, 1945). All of the males in each of two flocks were then permitted to remain together with the same hens used for testing. A summary of some of the data

is given in Table 1. Although differences in sexual drive between some of the males are suggested in the summary chart, there is also an indication that social rank influenced performance when in competition for mates. Of particular interest at this point is the case of Male III who was psychologically castrated in the presence of his social superiors. He also failed to tread the hens when all three of his ranking males were removed temporarily, but did so readily when placed into a pen with strange hens. It should be added that the sex ratio was one male to two females, whereas a ratio of 1-12 is less competitive among Leghorns. However, when a more reasonable sex ratio was used (Guhl & Warren, 1946) similar

TABLE 1

MEAN RATES PER HOUR OF SEXUAL ACTIVITY OF FOUR MALES IN EACH OF TWO FLOCKS TESTED SINGLY WITH THE HENS AND WHEN ALL FOUR MALES WERE IN THE PEN

Peck order	Courting		Treading	
	Singly	4 males	Singly	4 males
I	8.2	8.8	3.1	3.2
II	8.2	7.1	4.6	1.3
IV	7.2	6.4	2.5	2.5
III	8.7	2.5	3.3	0.0
V	24.3	11.3	9.5	1.6
VI	42.1	9.6	8.4	1.4
VII	21.1	3.2	7.1	0.2
VIII	41.3	10.0	9.2	1.1

Note.—From Guhl, Collias, and Allee (1945).

results were obtained. This extension of the study showed that the highest ranking male may also sire the most chicks in this polygamous species.

It is of further interest that the passive dominance of males over females is associated with normal mating. Young cockerels will fail to mate with older hens and are dominated by them. Since estrogen may induce sexual behavior in capons without any increased aggressiveness over the castrate condition, such males were used to determine the importance of heterosexual dominance (Guhl, 1949). Treated capons mated quite normally with females over which they were able to rank but were driven by those that held higher status in the heterosexual peck order formed prior to treatment. Thus social status may facilitate or inhibit sexual performance. Furthermore, it has been shown that the relative acquaintance between the sexes influences the frequency and type of sexual behavior patterns displayed by both sexes (Guhl, 1961).

*Social Status of Females.* Negative correlations have been found between the social rank of hens and the frequen-

cies of submitting to a male and of being mated (Guhl, Collias, & Allee, 1945). Hens may crouch in a submissive posture when strongly dominated by certain of their superior penmates, and pseudomatings between hens have been reported under such circumstances (Guhl, 1948). The sexual crouch displayed to a male is essentially similar but may be of a higher intensity. From such observations one might assume hens that rank low and are in the habit of submitting may display the receptive crouch more readily than high ranking females that are more inclined to domination. Such a hypothesis can be tested by altering the relative intensities of these habits.

The peck orders of three different flocks of approximately 30 hens each were determined (Guhl, 1950). Three or more males were caged within the pen and released singly for short daily periods. The number of crouchings observed for the hens in the top, middle, and bottom thirds of the social order was recorded. After several weeks of observation each flock was divided (subflocked) according to three levels in the peck order. This procedure did not alter the peck rights between the hens remaining in contact. Levels of domination and subordination were altered because the top third (as a group) now had two thirds fewer birds to dominate, the bottom birds had two thirds fewer dominating them, whereas the habits of the middle third were altered about equally between dominating and being dominated. The same males were then rotated through each of these respective subflocks. Subflocking occurred when the laying cycle was at its peak as an attempt to balance the endocrine relationships.

The results for one flock are indicated in Figure 4. Under the conditions of subflocking a total reduction in rates

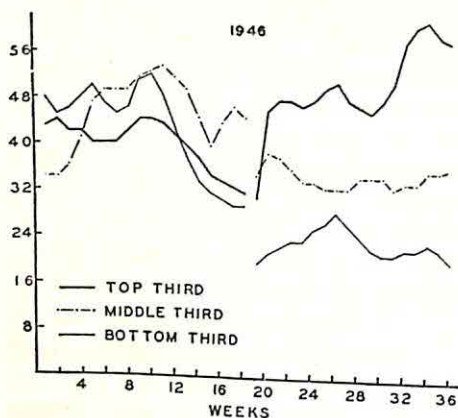


FIG. 4. The effects of habits of domination and submission on sexual receptivity as shown by subflocking that changed the intensity of domination in highest ranking hens and of submission by hens in lowest social level. (After subflocking by three levels in the peck order the dominating females from the top third became more receptive. See Guhl, 1950.)

of crouching was expected because only one male was used per small flock per day, whereas a sequence of three or more males was used in the large flock. In this manner the sex ratio remained essentially the same. However, in each test, the top third hens showed an increase in crouching response during subflocking, while the receptivity in the middle and bottom thirds decreased. No estimates were made of the sex drive of individual hens. Not all of the comparisons between thirds (on a week by week basis) within flocks were significant. Nevertheless, it may be concluded that the habit of dominating in females interferes with mating.

#### SOCIAL STRESS AND THE ENDOCRINE SYSTEM

Although the peck order is an adjustment to the presence of penmates, individuals at the lowest levels of the social order may be under stressors related to their rank, especially if situations become highly competitive. Since these birds are usually confined and under

high population density for economic reasons, the effects of social stress are now receiving some attention. According to Selye (1947) animals may adapt to various forms of stress by the increased activity of the pituitary-adrenal cortex axis, which becomes evident by the hypertrophy of the adrenals.

It has been shown that crowding, that is, the number of square feet per bird, results in an increase in the adrenal weight in hens (Siegel, 1959). However, there was no evidence of an acute hypertrophy of the cells or lipoidal depletion and the reaction was well within the birds' adaptive ability. Similar results were obtained with males (Siegel, 1960). When social stress was introduced experimentally among males (Siegel & Siegel, 1961) all individuals responded with increased adrenal weight when compared with controls in individual cages. The left adrenal, which is more sensitive than the right in chickens, was heavier in cockerels which ranked low in the social order. Within the limits of the stress to which they were subjected, the males remained in the adaptive phase. It was postulated that these birds may be less liable to psychological stimulation than mammals.

#### REFERENCES

- ALLEE, W. C., & COLLIAS, N. E. The influence of estradiol on the social organization of flocks of hens. *Endocrinology*, 1940, 27, 87-94.
- ALLEE, W. C., COLLIAS, N. E., & LUTHERMAN, C. Z. Modification of the social order in flocks of hens by the injection of testosterone propionate. *Physiol. Zool.*, 1939, 12, 412-440.
- ALLEE, W. C., & FOREMAN, D. Effects of an androgen on dominance and subordination in six common breeds of *Gallus gallus*. *Physiol. Zool.*, 1955, 28, 89-115.
- BEACH, F. A. Evolutionary aspects of psychoneuroendocrinology. In A. Roe & G. G. Simpson (Eds.), *Behavior and evolution*. New Haven: Yale Univer. Press, 1958. Pp. 81-102.
- CARPENTER, C. R. Characteristics of social

- behavior in non-human primates. *Trans. N. Y. Acad. Sci. Ser. II*, 1942, 4, 248-258.
- GUHL, A. M. Social discrimination in small flocks of the common domestic fowl. *J. comp. Psychol.*, 1942, 34, 127-148.
- GUHL, A. M. Unisexual mating in a flock of White Leghorn hens. *Trans. Kan. Acad. Sci.*, 1948, 51, 107-111.
- GUHL, A. M. Heterosexual dominance and mating behavior in chickens. *Behaviour*, 1949, 2, 106-120.
- GUHL, A. M. Social dominance and receptivity in the domestic fowl. *Physiol. Zool.*, 1950, 23, 361-366.
- GUHL, A. M. The development of social organization in the domestic chick. *Anim. Behav.*, 1958, 6, 92-111.
- GUHL, A. M. The effects of acquaintance between the sexes on sexual behavior in White Leghorns. *Poult. Sci.*, 1961, 40, 10-21.
- GUHL, A. M. The behaviour of chickens. In E. S. E. Hafez (Ed.), *The behaviour of domestic animals*. Baltimore, Md.: Williams & Wilkins, 1962. Ch. 17. (a)
- GUHL, A. M. The social environment and behaviour. In E. S. E. Hafez (Ed.), *The behaviour of domestic animals*. Baltimore, Md.: Williams & Wilkins, 1962. Ch. 5. (b)
- GUHL, A. M., & ALLEE, W. C. Some measurable effects of social organization in flocks of hens. *Physiol. Zool.*, 1944, 17, 320-347.
- GUHL, A. M., COLLIER, N. E., & ALLEE, W. C. Mating behavior and the social hierarchy in small flocks of White Leghorns. *Physiol. Zool.*, 1945, 18, 365-390.
- GUHL, A. M., CRAIG, J. V., & MUELLER, C. D. Selective breeding for aggressiveness in chickens. *Poult. Sci.*, 1960, 39, 970-980.
- GUHL, A. M., & WARREN, D. C. Number of offspring sired by cockerels related to social dominance in chickens. *Poult. Sci.*, 1946, 25, 460-472.
- MASLOW, A. H. Dominance quality and social behavior in infrahuman primates. *J. soc. Psychol.*, 1934, 11, 313-324. (a)
- MASLOW, A. H. Dominance and social behavior in monkeys. *Psychol. Bull.*, 1934, 31, 688. (b)
- SELYE, H. The general-adaptive-syndrome and the diseases of adaptation. In, *Textbook of endocrinology*. Montreal: Montreal University, 1947. Pp. 837-866.
- SIEGEL, H. S. The relation between crowding and weight of adrenal glands in chickens. *Ecology*, 1959, 40, 495-498.
- SIEGEL, H. S. Effect of population density on the pituitary-adrenal cortical axis of cockerels. *Poult. Sci.*, 1960, 39, 500-510.
- SIEGEL, H. S., & SIEGEL, P. B. The relationship of social competition with endocrine weights and activity in male chickens. *Anim. Behav.*, 1961, 9, 151-158.
- SIEGEL, P. B., & HURST, D. C. Social interactions among females in dubbed and undubbed flocks. *Poult. Sci.*, 1962, 41, 141-145.
- WILLIAMS, C., & MCGIBBON, W. H. An analysis of the peck-order of the female domestic fowl, *Gallus domesticus*. *Poult. Sci.*, 1956, 35, 969-976.

(Received November 28, 1962)

## RESPONSE PREFERENCES:

### A REVIEW OF SOME RELEVANT LITERATURE

G. S. TUNE<sup>1</sup>

*School of Public Health, Harvard University*

A review of literature giving evidence of response preferences, mainly in human Ss, is undertaken. The preferences are taken from experimental work in early psychophysical data, subjective estimates of "chance" sequences, response mechanisms at the threshold, "gambling" situations, probability learning, and the influence of instructions on performance in probability learning. Explanations of response preferences are reviewed and the relationships between subjective uncertainty and preferences are explored with an attempt to give cohesion to a diverse body of experimental evidence.

Fernberger (1913, 1920, 1930) was probably the first investigator to stimulate interest in the "contrast effect," a phenomenon which showed up as long sequences of nonrandom responses which distorted the psychophysical function and affected the position of the point of subjective equality. Turner (1931) confirmed Fernberger's finding that subjects tended to alternate their responses from response category to category and tried to explain this as a change in the effective value of the stimuli. Further experiments showed that the phenomenon was unaffected by the use of stimuli of equal physical value (Arons & Irwin, 1932), the use of long inter-stimulus intervals (Preston, 1936a, 1936b, 1936c), or the use of cross-modal judgment situations. In all cases the predominant pattern of response was an alternation.

Goodfellow's (1938) examination of the Zenith Radio Foundation experiments in telepathy led him to reject the telepathic interpretation of the results in favor of one in terms of subjective preferences for nonsymmetrical patterns

of response. He later suggested other factors which could affect the response choice (Gault & Goodfellow, 1940) and proposed that the subjective concept of "chance," which was different from the physical notion of the concept, would be important because subjects operating in accordance with this would be unable to generate a random series of responses. Skinner's (1942) explanation was somewhat similar. He described alternations as a "tendency" operating within the subjective idea of chance where an alternation of responses would diminish the tendency and a repetition would strengthen it.

This early work on the "contrast effect" was the historical root of much of the contemporary speculation on the stochastic nature of behavior. The findings had implied that the phenomenon was of central rather than peripheral origin (Irwin & Preston, 1937) and that long sequences of responses were likely to be statistically interdependent in a complex manner (Irwin & Preston, 1937; Preston, 1936b), two implications which have contemporary interest.

#### *Subjective Estimates of "Chance" Sequences*

Reichenbach's (1949) suggestion, that mathematically naive subjects

<sup>1</sup> My thanks are due to the Department of Scientific and Industrial Research of Great Britain which provided the financial support for this work and to Stephen Griew under whose guidance this work was undertaken at the University of Bristol.

would be unable to produce a random series of selections from a limited number of alternatives, has been explored experimentally. Number preferences, "gambling" situations, and successive judgments where no actual judgments could be made have all been used.

Subjects with some training in mathematics and probability theory were regarded as mathematically sophisticated and were compared with subjects without such training by Chapanis (1953). Subjects wrote a total of 2,250 numbers between 0 and 9 inclusive, in any order at their own pace. Subjective preferences for certain numbers were found but there was little agreement in their preferences. Repetitive pairs and triplets were avoided by unsophisticated subjects, while sequences of different numbers were preferred. Decreasing combinations were highly preferred but increasing combinations were not. Sequences of six, seven, and even eight numbers were sometimes repeated at widely separated intervals. An autocorrelation showed that there was an association between the  $n$ th and the  $n+x$ th digit, which was significant when they were separated by as many as four intervening digits. Sophisticated subjects tended to be more nearly random than the naive group.

A similar task was set by Lincoln and Alexander (1955). Eight discs were arranged to form either a square or a circle and subjects were instructed to produce a random selection by touching the metal discs with a stylus or calling out code names printed on paper discs. It was revealed that subjects did not select a random sample because of preferences for certain sequences of responses. The preferences were influenced by the shape of the figure used and the method of response.

Reichenbach's claim (1949), that unsophisticated subjects would be unable

to generate a random series because of insufficient runs of the same element, was specifically examined by Ross (1955). In this, subjects tried to produce a random series by stamping cards taken from a file with symbols X or O. The files were arranged so that the proportions of each symbol varied from condition to condition. Bakan (1960) challenged Ross' claim that the response series were random. Bakan obtained runs of responses from his subjects and compared the two halves of each set of data from each subject. Triplets of response correlations significant at the 1% level were found. Preferences for one alternative or the other ("heads" or "tails") gave  $r=.74$ , significant at the 1% level. On a chance expectation over 300 calls, 151 runs might be anticipated; in Bakan's experiment the mean number found was 176, 62 of the 70 subjects produced over 151 runs. A further analysis of the triplets of responses (e.g., HHH, HHT, HTH, etc.) showed that the homogeneous patterns (HHH and TTT) accounted for 16% of the total responses, whereas the alternative patterns (HHT, TTH, THH, HTT, HTH, and THT) made up 84%, which was 9% greater than would be expected if responses were distributed equally over all eight possible triplets. The symmetrical patterns (HTH and THT) made up 36% of the total and the asymmetrical (HHT, TTH, THH and HTT) 48%. Assuming that all patterns would be used equally often, these figures should be 25% for the former and 50% for the latter, indicating that the symmetrical patterns are over-used by 11%. These results supported Reichenbach's claim and since the examination of the response patterns was very thorough it seemed difficult to account for Ross's findings, unless they were explainable in terms of the type of statistic used.

The results of these experiments are fairly clear. Subjects did not appear able to give a random selection from groups of two or more alternatives, primarily because of preferences for certain patterns of response. Alternative response patterns and nonsymmetrical patterns seem to be preferred to long runs of homogeneous responses.

### *Response Mechanisms at the Thresholds*

Senders and Sowards (1952) presented subjects with a long series of weak flashes of light at a constant intensity, and a tone of constant volume, and required them to report when the two stimuli occurred together and when they occurred separately. Instructions led the subjects to believe that they were making discriminations, but in fact the stimuli were always presented together. Using an information measure developed by Newman (1951), Senders and Sowards (1952) found that subjects preferred long series of identical responses, a preference which decreased randomness and showed a tendency towards first-order redundancy. The authors suggested that the repetition of a response was easier than alternation, but the main conclusion was that when the "discrimination task" was impossible, stimulus determination of responses was not possible, and they seemed to become response determined.

Similar results in favor of response determination were found by Tune (1963). His subjects were required to state whether the gap in a Landolt-C was in one or other of two positions. He found that further away from the gap the degree of nonrandomness in the response series, as determined by the McGill (1954) statistic, increased. Further, responses became more response determined in that the relationship between the previous response and the response increased. Contrary to Senders

and Sowards (1952) and Howarth and Bulmer (1956), however, the predominant pattern of response was an alternation. The tendency suggested by Skinner (1942) was clearly shown.

The assumption made in the psychophysical methods, that responses are statistically independent of each other, was called into doubt by Senders (1953), Senders and Sowards (1952), Wertheimer (1953), and Verplanck, Collier, and Cotton (1952).

The general procedure followed by the authors was to ascertain their subjects' threshold by the method of limits and then embody this value in long series of stimuli. Subjects reported whether or not this stimulus was seen. Howarth and Bulmer (1956) presented subjects with stimuli of threshold intensity at 4-second intervals. Reports of "seen" or "not seen" were taken and the time-ordered responses were examined. Autocorrelations showed that the association between a response and subsequent responses disappeared very rapidly being almost completely gone at a lag of three. It had been suggested by Oldfield (1955, 1956) that fluctuations in visual sensitivity could account for such relationships as this, but Howarth and Bulmer (1956) were able to discount the suggestion. It was implied by the rationale of the experiment concerned with Oldfield's (1955) claim, that the predominating pattern of response was a repetition of the same response in long runs. This was contrary to what Tune (1963) found.

McGill (1957) first required his subjects to identify four tones, after which the four tones were presented in random order at a constant intensity. Difficulty in discriminating the tones was introduced by varying the level of background noise. The author reported that where the signal to noise ratio was below the average threshold, the response

series became nonrandom. This effect was greatest where only the background masking noise was present; it gradually declined as more intense tones were introduced.

These experiments showed that where discriminations were made on the basis of adequate objective evidence, then a random response series could result. However, as the discriminations became more difficult or stimuli could not be discriminated response sequences tended to exhibit a considerable degree of nonrandomness. The nature of this nonrandomness was a source of dispute, tendencies towards first- and second-order redundancy having both been found. Conklin (1954) drew a useful distinction between a judgment "set" and a guessing game "set." Assuming that suprathreshold responses to an auditory discrimination task were not significantly related he claimed that response independence was reasonable. He did not, however, publish details of the difficulty of discriminating the tones used in his experiment.

In a further attempt to justify the proposition that responses can at least in part be determined by stimulus events, Conklin and Sampson (1956) carried out a weight-lifting experiment with weights which differed from each other by as much as 64 grams or were of equal weight. They asked subjects to make the customary judgments in such a task and found that when the discrimination was difficult or impossible, then response determination was at a maximum.

Wertheimer (1953, 1954, 1955) found reliable shifts in both visual and auditory thresholds and explained them by saying that the successive samplings of thresholds were not independent. The day-to-day fluctuations were greater than the variations within 1 hour and the visual and auditory thresholds were

correlated, but because of the possibility of biased sampling, Wertheimer avoided speculating about a "central sensitivity factor."

The experiments by Day (1956a, 1956b, 1957) were remarkable in that they yielded contrary results: random and nonrandom responses at the threshold. In one experiment (Day, 1956a), subjects detected a 1 jnd in a tone previously determined to be at the 50% threshold. With interstimulus intervals of 1 and 9 seconds, Day found consistently random serial-threshold behavior. A similar set of experiments (Day, 1956b), with interstimulus intervals from 1.6 to 10.6 seconds, showed a decline in the amount of nonrandomness for intersignal intervals of longer than several seconds, but some subjects showed marked, if patchy, nonrandomness, at the longer stimulus intervals. Using the visual threshold (Day, 1957), the same author made experiments similar to those of Howarth and Bulmer (1956) with interstimulus intervals of 2, 4, 16, and 32 seconds. Nonrandomness of responses was found to be significantly less at the longer stimulus intervals. The shorter stimulus intervals showed patterning of both a first- and second-order type of redundancy, but there was no evidence of consistency from subject to subject in the way that the responses were organized.

Day (1956b) speculated that fatigue, boredom, wandering attention, physical discomfort, and attitudes toward the experiment introduced a certain amount of organization into the response series which may have accounted for the apparent periodicity of nonrandomness in his experiments. It is notable also that such a view had been implied by Welford (1953).

The most intensive investigations into the phenomenon under discussion were made by Verplanck and his col-

leagues who undertook an extensive examination of the response mechanisms at the visual threshold. Verplanck, Collier, and Cotton (1952) examined subjects' time-ordered responses to a weak flash of light previously determined to be at the 50% visual threshold. Two conditions were used in which stimuli were paced and self-paced. Autocorrelations were calculated and it was reported that even with a lag of nine responses there was some degree of association. The correlation, as in Howarth and Bulmer's (1956) work, was a monotonically decreasing function of the lag.

It was possible that determining the visual threshold by the ascending and descending method of limits could produce a "set" or "expectancy" for long runs of the same response, an effect which might have been particularly prevalent in well-trained subjects. In order to examine this, Verplanck, Cotton, and Collier (1953) presented flashes of light of various intensities in random order. This procedure would have eliminated any predisposition subjects had for responding repeatedly with the same category. It was found that a group given this procedure did not produce different results from a group which experienced the usual intensity-ordered procedure; both groups showed serial dependencies in the form of long runs of the same response. As the experiment proceeded, however, the threshold was raised in that there was a decreased frequency of seeing the flash of light. This was later confirmed by Bakan (1955) who reported that as a visual vigilance task proceeded the threshold rose. Verplanck, Cotton, and Collier (1953) concluded that although a rise in the threshold might produce nonrandom series of responses there was evidence that response-to-response dependencies could account for it at least in part. Brackmann and Collier (1958)

found that the visual threshold determined from an ascending series of stimuli was lower than that determined from a descending series; the difference between these was an increasing function of the size of the step interval, the variability of the psychophysical function being directly related to the size of this interval. This was interpreted as a product of an interaction of two tendencies: first, to alternate responses which was a function of the length of the preceding sequence of responses of the same category; second, a tendency to repeat responses which was a decreasing function of the size of the step interval and of the preceding response.

Examining response chains at the visual threshold, Collier and Verplanck (1958) showed that if stimuli of high (suprathreshold) intensity or low (subthreshold) intensity were interpolated into a run of stimuli of invariant brightness, the sequential characteristics of the responses could be disturbed. Obviously the suprathreshold stimuli should force a response "seen" and the subthreshold ones the response "not seen." It was shown that the magnitude of interresponse association following these interpolated stimuli was less than that between responses taken directly from the invariant series. This finding was of considerable importance since it indicated that actual stimulus differences were likely to be ignored in favor of what might be called a preoccupation with responses rather than stimuli.

Further evidence of this tendency was found by Verplanck and Blough (1958). They reported that responses were not statistically independent, runs of the same responses being a common feature which produced this effect. Response-to-response dependencies, however, did not account for all the nonrandomness and fluctuating "sensitivity" factor was a possible contributing course.

Collier (1954) found no reduction in

the amount of association between successive responses in binocular as compared with monocular conditions and concluded that the nonrandomness was at least, in part, of nonperipheral origin adding force to Irwin and Preston's (1937) conclusions.

In discussions of the possible mechanisms governing response behavior, Brackmann and Collier (1958) and Collier and Verplanck (1958) noted that the tendency to alternate was a feature of experiments which required the subject to guess; this was typified as a simple response-to-response dependency. The tendency to repeat responses was said to be more complex in that it was a response-outcome dependence and a function of two variables rather than simply one. The situation was probably not as simple as this, however, McGill (1954) demonstrated that 28% of the response information was due to an association between the stimulus and the response and the preceding response, a half of this (14%) being a result of interaction between the stimulus and the preceding response. This indicated that a response-previous-response dependency existed in forced-choice tasks, but as Collier and Verplanck (1958) noted, stimulus similarity was an important factor. The more difficult it was to discriminate between several stimuli, the greater was the likelihood of there being response-to-response dependency. Crossman (1955) pointed out the need for an objective measure of discriminability and, from a communication theory background, developed a "confusion function" which gave such a measure. Verplanck and his co-workers have not used this measure although it has been available for some time.

Considerations of response mechanisms at the visual threshold only, led Collier and Verplanck (1958) to suggest four empirical constraints on any theoretical model of response behavior.

They said that the magnitude of association was related to the similarity between successive stimuli, was related to the general level of stimulus intensities, was a decreasing function of the inter-trial interval, and may be related to the number of categories of judgment.

This was particularly a criticism of the model originated by Weiss, Coleman, and Green (1955) which appeared to oversimplify the elaborate system of dependencies uncovered by Verplanck and his co-workers.

### *Response Preferences in "Gambling" Situations*

Bakan (1960) reported that 80% of his subjects called "heads" at the first toss of a coin. Goodfellow (1940) gave a similar figure, also showing that this preference had fallen to 57% on the second toss and 44% on the third. There is a considerable body of research in existence on this topic, in particular that of Cohen (1960) and other members of the psychology department at Manchester University, England. Only a small sample of the more relevant papers will be discussed here.

Gambling behavior has been studied by Edwards (1953, 1954, 1956) who found that there were probability preferences. These preferences were shown mostly when the odds were equal. A further finding of relevance was that gamblers tended to stay with a winning strategy but when losing were more ready to change.

No relationship has been reported between intelligence and subjective estimates of probability, but Dale (1958) added the qualification that intelligence determined not the hunches the man had, but whether he chose to be guided by them. Whatever the intelligence of the gambler some patterning in the betting behavior was apparent. Dale (1959) further demonstrated the tendency of subjects to overestimate the value of

events with a low probability and a high gain, and to underestimate the value of events with a high probability and a low gain, in an experiment with "long odds" and "short odds." A relationship was found between subjective probabilities and subsequent behavior. This was not simple and there was reason to believe that some of the subjects were relatively sophisticated and were likely to make a bet and then change their opinions about the wisdom of their choice.

In a situation designed to look like a gambling task (selecting draws on a pseudo-Pakistani football coupon), it was shown that the selection of items demonstrated a kind of sequential dependence (Dale, 1958) due to "bunching" and "spacing" effects which were too regular to have occurred by chance. It appeared that selections were made in accordance with some nonrandom system.

Although gambling behavior is very modifiable, it seems to be strongly influenced by the ideas that gamblers hold about the probabilities of the events they deal with.

#### *Probability-Learning Experiments*

In a gambling situation subjects are required to place bets on certain options after considering the consequences of the outcomes of these options, along with the potential rewards or losses. Much the same can be done in probability-learning experiments, but often the rewards are not financial; the subject being given merely the knowledge that his choice was correct or otherwise.

There are basically two types of probability-learning experiments. In both, series of events are presented according to some stochastic process with a variety of available protocols. In the first type of experiment, the subjects are presented with a run of events and are then asked to estimate with what

frequency each event occurred. In the second type of situation, subjects are required to predict which of the events will next occur.

Hornseth (quoted by Hake, 1955) used an experimental arrangement of the former type. Subjects observed a sequence of two lights flashing: on some proportion of the sequence both lights were illuminated; on the remainder, only one—the proportions were varied from condition to condition. Plotting the relative frequency of occurrences as estimated by the subjects against the objective frequency of occurrences, a typical phi-gamma function was found attributable to an overestimation of the high and an underestimation of the low frequencies.

Similar inaccuracies were found by Attneave (1953) when subjects estimated the frequency with which letters occurred in English. This was taken as evidence that even long experience, of a stochastic process, need not effect the subjective estimates of relative frequency of occurrence of the events. It is important to note that, in Attneave's situation there were 26 events, in Hornseth's only two; nevertheless, there is some indication that Collier and Verplanck (1958) were right in placing a fourth empirical constraint on any stochastic model of choice behavior.

Using the second type of experimental procedure with a variety of unbalanced schedules with two events, Grant, Hake, and Hornseth (1951) found that whatever the structure of the event series, subjects initially predicted each event with equal frequency; in other words, they seemed to hold an *equiprobable* hypothesis. This appeared in many experiments of this type (Grant, 1953; Jarvik, 1951; Lincoln & Alexander, 1955; Nicks, 1959; Senders & Sowards, 1952; Tune, 1963). In experiments where the series of events departed from

the equiprobable state (Grant, Hake, & Hornseth, 1951), the response series generated by the subjects gradually abandoned the equiprobable hypothesis and approximated the statistical structure of the event series.

Jarvik (1951) presented subjects with lists of two words arranged in random order asking them to guess which word would next occur and then informing them which word did actually occur. One word held preponderance over the other in a variety of ratios (60%, 67%, and 75%). Typical probability-learning curves were produced, that is, in the 75% condition, the asymptotes were reached more quickly than the asymptotes in the 67% condition and these again more rapidly than in the 60% condition. It was noted that once the asymptotes were reached, the curves tended to decline somewhat. Brunswik and Herma (1951) described similar curves as having "a slow initial rise followed by a slow 'paradoxical' decline without reaching statistical significance." The sequences of responses showed a positive recency effect; that is, long runs of the same response.

This type of study was enlarged upon by Goodnow and Postman (1955). It was found that in a 100:0 (two-choice) condition the asymptotes were approximated more quickly than in any other and a decline in performance of the type mentioned above was thereafter evident. Goodnow and Pettigrew (1955, 1956) and Goodnow, Rubinstein, and Lubin (1960) made further studies along the same lines but with the difference that events were first presented with one frequency balance in a specific pattern, were then changed to a random frequency distribution, and subsequently to the initial one or yet another. Subjects were thus required to learn an initial-response strategy, were then forced to abandon this because events were only

randomly predictable, and then had to assume the initial or another response strategy.

The implication of the results of these experiments was that any disposition to look for sequences of events was an advantage over having no disposition. Goodnow and Pettigrew (1955) showed similar results on the same kind of experimental design with highly redundant event sequences.

From communication theory it can be shown that in a two-choice situation making events equally probable maximizes average amount of information. The maximum can be reduced either by unbalancing the probabilities of occurrence or introducing sequential dependencies into the event series or both. This is what Goodnow, Rubinstein, and Lubin (1960) did. In agreement with previous findings, they reported that subjects who had experienced structured runs in the first set of events most like the run structures in the second set of events, showed the greatest speed in adapting from one to another. These authors also pointed out the need for an independent measure of discriminability, which is much the same as that which Grant (1953) was looking for, and drew the conclusion that an important factor in the process of discriminating sequences of events was the subjects' interest in his own responses. This is the response-to-response dependency found by other authors and indicated that even when event series were presented at suprathreshold values, subjects tended to be influenced by responses which they made earlier in a response series. This was perhaps inevitable because it might be reasonably assumed that subjects were testing out a variety of response strategies, which required a certain level of disproof before they were discarded.

Edwards (1956) specified a way in

which different types of response strategies, or "hypotheses," might operate. He suggested that:

Big hypotheses define classes of small hypotheses. Small hypotheses, which usually specify sequences of choices, cannot be correct, and so are constantly being tried, abandoned, and replaced with other small hypotheses [p. 187].

A big hypothesis could be that the events have a general frequency distribution. A small hypothesis, for example, could be that certain transitional probabilities applied within the context of the big hypothesis, and since there are very many of the former, it is likely that subjects would test a good many before abandoning one big hypothesis in favor of another.

In a similar manner Nicks (1959) speculated about the way in which subjects approached probability-learning situations. He noted that subjects initially appeared to hold the equiprobable hypothesis which was subsequently modified, if it was proved to be inappropriate. Nicks (1959) also made the point that his subjects seemed to bring along to the situation certain a priori expectations concerning the structure of the sequence. These expectations were like the small hypotheses mentioned by Edwards (1956) and indicated that while holding a very general equiprobable hypothesis, subjects seemed to have preferences for patterns of events possibly reflecting their ideas of "chance."

Further evidence that patterning of events was of interest to subjects in probability-learning tasks was produced by Komorita (1958). He found that equivalent methods of collecting data yielded equivalent results in that the learning curves for equivalent methods were not significantly different. An exception to this generalization was the finding that subjects did not seem to respond to single signal occurrences as they did to patterns of stimulus occur-

rences even though the probabilities of the stimulus events were identical. This finding indicated that the shapes of the learning curves using Komorita's technique were not the same as the curves obtained by Humphreys (1939) or Jarvik (1951). Although the asymptotes were the same, the rates of learning differed, implying that patterned events produced a different rate of learning from that of learning single events.

Although there have been some experiments using multichoice tasks (Detambel, 1955; Gardner, 1957; Neimark, 1956), the majority of theories of decision making have been derived from a consideration of two-choice tasks (Anderson, 1955; Bush & Mosteller, 1955; Detambel, 1955; Estes, 1950, 1955). Some authors have claimed a high degree of correspondence between the theories they propose and the empirical findings; but none of these, so far, have taken into account all the constraints proposed by Collier and Verplanck (1958).

#### *Influence of Instructions on Performance in Probability-Learning Tasks*

"Recency effects" (Jarvik, 1951) are patternings of responses or sequential dependencies which exhibit tendencies to alternate from category to category (negative recency effects, which are 100% redundant at the second order) and a tendency to repeat responses (positive recency effects, which are 100% redundant at the first order). It has been suggested, however, that these recency effects are an artifact of instructions given to subjects.

Goodnow (1955) hypothesized that a task said to involve chance would produce recency effects but a similar two-choice task said to exclude chance would not. Two tasks were used: a gambling situation in which it was implied that chance was involved and a problem-solving task in which it was implied that no chance was involved. Negative recency

effects were found in the gambling task where, it is of interest to note, some subjects consistently chose the event which was most frequently rewarded as predicted by Flood's (1954) theory. This response pattern has been found rarely, the reason being (Hake, 1955), that subjects had a primary interest in predicting the infrequent event and this interest precluded the uniform prediction of the most frequent event. Furthermore, subjects were indifferent to eventual outcomes because of the simplicity of the probability-learning situation, and "reactive inhibition" may have reduced the frequency with which subjects responded to the most frequent event.

Van Der Meer (1960) argued that instructions suggesting that the sequence of events to be predicted was nonstationary would produce a higher asymptotic value of the predictions of the more frequent event than instructions that the process was stationary. He tested his argument in a two-choice situation and found that it was substantially correct. The author showed the importance of this finding for mathematical learning theories of the Bush-Mosteller and Estes-Burke type, which were based on the assumption that positive recency effects would predominate. This tendency has been widely reported throughout the experimental literature (Howerth & Bulmer, 1956; Senders, 1953; Senders & Sowards, 1952), but there is considerable evidence that negative recency effects occur just as often (Brown & Overall, 1959; Fernberger, 1930; Grant et al., 1951; Jarvik, 1951). Van Der Meer (1960) proposed that the differences in asymptotic values found between the two conditions of instructions were mainly determined by differences in motivation which made the utility of a correct prediction greater in the case of instructions implying that the process was stationary.

Similar results were found by Hyman and Jenkin (1956). In a two-choice experiment it was reported that subjects led to believe that events were entirely predictable tended to alternate an incorrect response more than subjects told that the event series was random.

The experiment by Senders and Sowards (1952) mentioned above produced perhaps the most startling results showing the influence of instructions on choice behavior. If subjects were instructed that events would occur approximately 75% of the time together and 25% of the time separately, their responses showed this inequality. Actual stimulus information was ignored in spite of the fact that both the flash of light and the tone were presented at suprathreshold levels.

It appears from this and other evidence quoted that instructions are of great importance in choice situations because they can bias the results in a number of ways. Brown and Overall (1959) sounded a warning, that care should be taken in applying learning theories to experimental data especially where there was any likelihood of recency effects of a type contrary to those assumed. Similar warnings have been made by or implied by Overall and Brown (1957), Edwards (1961), Feldman (1959), and Lindman and Edwards (1961). The paper by the latter workers reported the interesting finding that both positive and negative recency effects could be located in the same set of responses in a probability-learning situation, depending essentially on how long the task lasts.

#### *Grouping of Responses and Various Findings*

There is considerable experimental evidence to support the claim that human reaction time to an event is a function of the information content of the event (Brown, 1960; Crossman,

1953; Hick, 1952; Hyman, 1954). It is also known that the average information content of an event is reduced by making the series in which it is contained redundant in some way. Although "Hick's law" (Welford, 1961) has recently been called into doubt (Alluisi, Muller, & Fitts, 1957; Leonard, 1954; Mowbray, 1961; Mowbray & Rhoades, 1959), the final status of the law has not yet been determined.

In a probability-learning situation Hake and Hyman (1953) found that responses, examined as a function of preceding responses, showed that subjects responded to sequences of events including preceding predictions. Homogeneous runs of correctly predicted symbols were more "adequate" than others. This type of work was continued by Senders and Cohen (1955) who introduced sequential dependencies into the settings of pointers on dials on successive readings. They found that performance improved when measured in terms of information transmitted or accuracy. Redundancy can be regarded as a safeguard against error and misunderstanding in communication (Attneave, 1959), hence Senders and Cohen's results are not entirely unexpected. It has been well established that language is highly redundant (Newman & Gerstman, 1952) and in situations where accurate communication is of importance as in air traffic control, it is not surprising that the messages passed are more redundant than the language in general (Fritz & Grier, 1955).

Learning a language can be regarded as learning a complex system of events with their probabilities of occurrence. Bennett, Fitts, and Noble (1954) carried out a simplified version of this process. They used equally probable events which were generated according to a restricted random procedure, such that redundancies at various orders of estima-

tion were present. Since there were five events, the number of possible combinations was very large indeed. In spite of this, subjects learned digram sequences to a significant level, an additional finding being that the responses were more redundant than the stimuli because of pattern preferences. A second experiment demonstrated that subjects could learn both the preferred or nonpreferred sequences although initially the nonpreferred sequences presented difficulty. This might be interpreted to mean that initial preferences can be rapidly modified. Trigram sequences were not learned in general, probably because there were 50 trigram patterns which would require subjects to retain a considerable amount of sequential material.

It was suggested that learning engram, digram, and trigram patterns was inversely related to the total number of possible events. This would, so Bennett, Fitts, and Noble (1954) claimed, explain the failure of their subjects to show significant learning of the trigram patterns. A contributing cause might have been that the interstimulus interval of 4 seconds was so long that it was difficult for subjects to retain much of the information in immediate memory and it would also have militated against the tendency to "group" sequences together.

Grouping of responses and recognizing patterns was studied by Brown (1960). He wanted to know the most efficient way of grouping from 3 to 15 signals. The 3-choice, 7-choice and 15-choice situations were extracted from displays of lights which were presented singly or in combinations; combining sources allowed fewer lights to be used. At the 3 and 7-choice levels there was no significant difference between the rate of gain of information from the two types of display. At the 15-choice level a significantly greater amount of information

gain was reported from the single-source display, and reaction time against information in the display showed a linear relationship.

The importance of human ability to perceive and respond to patterns of stimuli is of interest mainly to research workers who speculate about the processes of information transmission, but it has implications for both the organization of behavior and more specifically for the phenomenon of anticipation in skilled performance. The literature reviewed does not allow definite conclusions to be made in all cases. Often the evidence presented from different sources appears contradictory, but where possible this has been explained as a variation in interpretation or a result of the statistical techniques used. In many instances there are gaps remaining in the experimental evidence which are a hindrance to any attempted generalizations, and it is here that extrapolation must be used but with caution.

#### *Some Explanations of Response Patterning*

Much work has been done on animals in the familiar Y or T maze, where the sequentially ordered list of choices of pathway exhibited the effects under discussion (Dennis, 1939; Dennis & Hennehan, 1939; Heathers, 1940; Hunter, 1914; MacGillivray & Stone, 1930; Tolman, 1925; Yoshioka, 1929). The behavior was often described as an "exploratory tendency" (Tolman, 1925) and there were many attempts to explain this tendency in terms of work decrement or fatigue (see Solomon, 1948, for a review of the literature). A further explanation was in terms of reactive inhibition (Zeaman & House, 1951).

Skinner's (1942) explanation of the contrast effect in terms of a subjective concept of "chance" allowed the possibility that such tendencies are innate.

At that time alternation tendencies predominated in experimental findings in both human and animal performance; since then, the positive recency effect has been found, along with demonstrations that such effects might be artifacts. The general explanation Skinner offered did not deny that both positive and negative recency effects were compatible with such a concept. Obviously chance could mean response sequences which showed alternations to some subjects, and to others, repetitions of the same response.

Solomon (1949) claimed that reactive inhibition could not explain the tendency to alternate. The main point of the argument was that reactive inhibition would dissipate in time and thus long interstimulus intervals would reduce the tendency. Glanzer (1953) offered a similar theory in terms of stimulus satiation. Bendig (1951) thought that Hull's reactive inhibition postulate implied something other than that interpreted by Solomon (1949). It predicted, for example, that rewarding or reinforcing a response, such as a guess, would have the effect of increasing the probability of a repetition of that response on the next elicitation, and that the effect would be greater the larger the number of rewarded responses. Bendig (1951) found this to be so and preferred not to abandon the reactive inhibition theory.

The production of nonrandom sequences of responses, by whatever means, led Senders and Sowards (1952) to speculate that since psychophysical methods assume the independence of successive responses, a good deal of the work done using psychophysical methods must be called into doubt. Conklin (1954) pointed out that in Senders' (1953) experiment, subjects could only guess. In a true psychophysical situation, there are supraliminal as well as subliminal

and threshold events; thus a judgmental set as opposed to a guessing game develops, especially where there are three categories of response. Conklin found that if real differences in stimulus values were used, then subjects could respond to them correctly. This implied that the type of experimental setup could suggest a method of response to a subject, or at least create a set to respond in a certain way.

Weiss, Coleman, and Green (1955) developed a mathematical model from the data in which subjects were required to bisect a perceived angular extent on the basis of kinesthetic cues. Subjects tended initially to drift into constant error, and long series of bisections yielded nonrandom response sequences as a result of positive-recency effect. A stochastic model for this type of response behavior was described as "a Markoff (first-order autoregressive) process." The implication of this was that in order to predict what a particular response was, only the previous response was necessary; earlier responses being of little use in that they did not add to the accuracy of the prediction. The authors concluded that there existed some kind of "kinesthetic trace," the duration of which was small. This particular model was carefully restricted to responses made in a kinesthetic judgment situation. The claim that remote past events were of no consequence in the making of judgments indicated that many small-range sequential dependencies were evident. In visual-threshold experiments, however, response sequences show much larger-range sequential dependencies, (Howarth & Bulmer, 1956). An interpretation of these two findings may well be in terms of "kinesthetic" and "visual" traces, the latter having greater duration.

The evidence suggests that, in general, as the distance between two responses in a response chain is increased, the

correlation diminishes. If decision processes are mediated by a central mechanism, it is hard to account for the lessening of sequential dependence in a kinesthetic judgment task as opposed to a task which involves visual, auditory, or tactual judgments. Apart from the tentative explanation offered above, however, a further point can be made. The extent of sequential dependencies in a response series may reflect the degree of uncertainty a subject experiences in making responses. If stimulus determination is impossible, then a guessing game set must be assumed and the sequential effects are entirely the result of response-previous-response or pre-previous-response influences. At the other extreme, where stimulus determination is entirely possible, then any sequential effects must be a reflection of sequential dependencies in the stimulus events. At points in between these extremes there is neither complete stimulus determination nor complete response determination. In this region responses may be partly determined by events in both the stimulus chain and in the response chain, a situation which exists in threshold experiments. Unfortunately most research has examined response sequences at the threshold in terms of sequential dependencies in the response chain only. McGill (1957) has demonstrated that both stimuli and previous responses can determine response behavior.

Since responses at threshold usually require a discrimination to be made, it follows that if the discriminations are difficult, then a decision, or response, or judgment must be made on the basis of less information than is necessary for an objective assessment. The magnitudes of just noticeable differences are different for each sensory mode, thus they require varying amounts of information for decisions to be made. It may be that the kinesthetic judgments required of

the subjects by Weiss, Coleman, and Green (1955) provided a dependency of the Markov type because the judgments were relatively easy to make; that is, there was sufficient information in the stimulus series to minimize the effect of response determination. In situations where discriminations or decisions are much harder to make, then more complex sequential dependencies are to be expected in the response chain.

It is easy to criticize mathematical models of the stochastic kind because they tend to oversimplify the situation. Howarth and Bulmer (1956), however, have indicated, and Bulmer and Howarth (1957) have demonstrated, that models of behavior which try to take into account long-range sequential effects become very complex, as anticipated by Arons and Irwin (1932). In their model of nonrandomness in threshold experiments (Bulmer & Howarth, 1957), it was assumed that sequential effects were comparatively short-lived. They also suggested that:

It is obviously good strategy (*for S*) to set his threshold low if the stimulus occurs often and high if it occurs seldom, since the chief danger in the first case will be to fail to see a stimulus which is there, and in the second case to see it when it is not there. A tendency to repeat his previous response will achieve this object of raising the average threshold when the stimulus is rare, and lowering it when the stimulus occurs frequently [p. 8; italics added].

In comparison, experiments where avoidance of repetitions seemed to be the prevalent pattern of response, it was suggested as reasonable that the test stimulus would be greater than the standard stimulus on about half of the occasions.

In such a case the best strategy would be to fix the threshold at zero and to leave it there. However, if there is any danger of a constant error resulting from the threshold being fixed some distance away from zero, alternation of responses becomes a good strategy, since it will have the effect of shifting the average threshold towards zero, provided, of course, that the

greater and less comparisons are presented equally often [p. 9].

If forced-choice judgments of the sort reported by Senders and Sowards (1952) can be regarded as comparisons then this explanation need not apply. Tune (1963) found alternation, but here there was the possibility of some objective evidence from which judgments could be made, whereas in Senders and Sowards (1952) and Senders (1953) there was not.

The complexity of the theoretician's task in building a mathematical model of choice behavior at the threshold, or otherwise, is evident. There are many variables in operation and there is a bewildering amount of experimental evidence emphasizing the importance of each. There are also a great many statistical techniques in use for detecting the existence and evaluating the range and strength of sequential dependencies. Statistics derived from communication theory such as those described by Newman (1951) and Attneave (1959) describe the sequential dependencies in sequences of events, but they cannot treat both stimulus and response events simultaneously or deal with interaction terms. The multivariate information transmission technique (McGill, 1954) and the symmetric uncertainty analysis (Garner, 1957) are perhaps the most powerful communication measures yet developed. Both are multidimensional—like the parametric analysis of variance—are capable of dealing with several variables simultaneously, and can give measures of interactions between variables. The applications of such statistics to the type of research discussed in this review are, it is felt, likely to yield the most important results.

#### REFERENCES

- ALLUISI, E. A., MULLER, P. F., & FITTS, P. M.  
An information analysis of verbal and motor responses in a forced-pace serial task. *J. exp. Psychol.*, 1957, 53, 153-158.
- ANDERSON, N. H. Effect of double stimulus

- events in a verbal conditioning situation. *Amer. Psychologist*, 1955, 10, 466.
- ARONS, L., & IRWIN, F. W. Equal weights and psychophysical judgments. *J. exp. Psychol.*, 1932, 15, 733-756.
- ATTNEAVE, F. Psychological probability as a function of experienced frequency. *J. exp. Psychol.*, 1953, 46, 81-86.
- ATTNEAVE, F. *Applications of information theory to psychology*. New York: Holt, 1959.
- BAKAN, P. Discrimination decrement as a function of time in prolonged vigil. *J. exp. Psychol.*, 1955, 50, 387-390.
- BAKAN, P. Response tendencies in attempts to generate random binary series. *Amer. J. Psychol.*, 1960, 73, 127-131.
- BENDIG, A. W. The effect of reinforcement on the alternation of guesses. *J. exp. Psychol.*, 1951, 41, 105-107.
- BENNETT, W. F., FITTS, P. M., & NOBLE, M. The learning of sequential dependencies. *J. exp. Psychol.*, 1954, 48, 303-312.
- BRACKMANN, J., & COLLIER, G. The dependence of probability of response on size of step interval in the method of limits. *J. exp. Psychol.*, 1958, 55, 423-428.
- BROWN, I. D. Many messages from few sources. *Ergonomics*, 1960, 3, 159-168.
- BROWN, W. L., & OVERALL, J. E. Implications of recency effects for probability learning theories. *J. gen. Psychol.*, 1959, 61, 243-251.
- BRUNSWIK, E., & HERMA, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, 41, 281-290.
- BULMER, M. G., & HOWARTH, C. I. A model of non-randomness in threshold experiments. *Air Ministry Flying Personnel Res. Comm., Great Britain*, 1957, No. 974(a).
- BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.
- CHAPANIS, A. Random number guessing behavior. *Amer. Psychologist*, 1953, 8, 332.
- COHEN, J. *Chance, skill and luck*. London: Penguin Books, 1960.
- COLLIER, G. W. Probability of response and inter-trial association as functions of monocular and binocular stimulation. *J. exp. Psychol.*, 1954, 47, 75-83.
- COLLIER, G. W., & VERPLANCK, W. S. Non-independence of successive responses at the visual threshold as a function of interpolated stimuli. *J. exp. Psychol.*, 1958, 55, 429-437.
- CONKLIN, J. E. Senders on response sequences. *Amer. J. Psychol.*, 1954, 67, 363-365.
- CONKLIN, J. E., & SAMPSON, R. Effect of stimulus determination on response independence at threshold. *Amer. J. Psychol.*, 1956, 69, 438-442.
- CROSSMAN, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice response. *Quart. J. exp. Psychol.*, 1953, 5, 41-51.
- CROSSMAN, E. R. F. W. The measurement of discriminability. *Quart. J. exp. Psychol.*, 1955, 7, 176-195.
- DALE, H. C. Subjective probability, gambling and intelligence. *Nature*, 1958, 181, 363-364.
- DALE, H. C. A priori probabilities in gambling. *Nature*, 1959, 183, 842-843.
- DAY, W. F. Randomness of threshold responses at long inter-stimulus intervals. *Percept. mot. Skills*, 1956, 6, 205-208. (a)
- DAY, W. F. Serial non-randomness in auditory differential thresholds as a function of inter-stimulus intervals. *Amer. J. Psychol.*, 1956, 69, 387-394. (b)
- DAY, W. F. A descriptive analysis of the non-randomness of serial threshold data. *Amer. J. Psychol.*, 1957, 70, 238-242.
- DENNIS, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, 28, 305-312.
- DENNIS, W., & HENNEMAN, R. H. The non-random character of initial maze behavior. *J. genet. Psychol.*, 1939, 40, 396-405.
- DETAMBEL, M. H. A test of a model for multiple-choice behavior. *J. exp. Psychol.*, 1955, 49, 97-104.
- EDWARDS, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, 66, 349-364.
- EDWARDS, W. The reliability of probability preferences. *Amer. J. Psychol.*, 1954, 67, 68-95.
- EDWARDS, W. Reward probability, amount and information as determiners of two-alternative decisions. *J. exp. Psychol.*, 1956, 52, 177-188.
- EDWARDS, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, 62, 385-390.
- ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 97-107.
- ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, 62, 145-154.
- FELDMAN, J. On the negative recency hypothesis in the prediction of a series of binary symbols. *Amer. J. Psychol.*, 1959, 72, 597-599.
- FERNBERGER, S. W. On the relation of the method of just perceptible differences and constant stimuli. *Psychol. Monogr.*, 1913, 14(4, Whole No. 61).
- FERNBERGER, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, 3, 126-150.
- FERNBERGER, S. W. The use of equality judg-

- ments in psychological procedures. *Psychol. Rev.*, 1930, 37, 107-112.
- FLOOD, M. M. On game-learning theory and some decision making experiments. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954. Pp. 139-158.
- FRITZ, L., & GRIER, G. W. Pragmatic communication, a study of information flow in air traffic control. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955. Pp. 232-243.
- GARDNER, R. A. Probability learning with two and three choices. *Amer. J. Psychol.*, 1957, 70, 174-185.
- GARNER, W. R. Symmetric uncertainty analysis and redundancy of printed English. *Proc. 15th Int. Congr. Psychol., Brussels*, 1957, 104-110.
- GAULT, R. H., & GOODFELLOW, L. D. Sources of error in psychophysical measurements. *J. gen. Psychol.*, 1940, 23, 197-200.
- GLANZER, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, 60, 257-268.
- GOODFELLOW, L. D. A psychological interpretation of the results of the Zenith radio experiments in telepathy. *J. exp. Psychol.*, 1938, 23, 601-632.
- GOODFELLOW, L. D. The human element in probability. *J. gen. Psychol.*, 1940, 23, 201-205.
- GOODNOW, J. J. Response sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, 68, 624-630.
- GOODNOW, J. J., & PETTIGREW, T. F. The effect of prior patterns of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, 49, 381-389.
- GOODNOW, J. J., & PETTIGREW, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, 51, 385-392.
- GOODNOW, J. J., & POSTMAN, L. Learning in a two-choice probability situation with a problem solving setting. *J. exp. Psychol.*, 1955, 49, 16-32.
- GOODNOW, J. J., RUBINSTEIN, I., & LUBIN, A. Response to changing patterns of events. *Amer. J. Psychol.*, 1960, 73, 56-67.
- GRANT, D. A. Information theory and the discrimination of sequences of events. In B. McMillan (Ed.), *Current trends in information theory*. Pittsburgh: Univer. Pittsburgh Press, 1953.
- GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with different percentages of reinforcement. *J. exp. Psychol.*, 1951, 42, 1-5.
- HAKE, H. W. The perception of frequency of occurrence and the development of expectancy in human experimental subjects. In H. Quastler (Ed.), *Information theory in psychology, problems and methods*. Glencoe, Ill.: Free Press, 1955. Pp. 257-277.
- HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, 45, 64-75.
- HEATHERS, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, 10, 359-380.
- HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, 4, 11-26.
- HOWARTH, C. I., & BULMER, M. G. Non-random sequences in visual threshold experiments. *Quart. J. exp. Psychol.*, 1956, 8, 163-171.
- HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, 25, 294-301.
- HUNTER, W. S. Auditory sensitivity in the white rat. *J. anim. Behav.*, 1914, 4, 215-222.
- HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1954, 45, 188-196.
- HYMAN, R., & JENKIN, N. W. Involvement and set as determinants of behavioral stereotypy. *Psychol. Rep., Monogr. Suppl.*, 1956, No. 3, 131-146.
- IRWIN, F. W., & PRESTON, M. G. Avoidance of repetition of judgments across sense modalities. *J. exp. Psychol.*, 1937, 21, 511-520.
- JARVIK, M. G. Probability learning and negative recency effect in serial anticipation of alternate symbols. *J. exp. Psychol.*, 1951, 41, 291-297.
- KOMORITA, S. S. Probability learning under equivalent data collecting methods. *J. exp. Psychol.*, 1958, 55, 115-120.
- LEONARD, J. A. The effect of partial advance information. *Med. Res. Council Rep., Appl. Psychol. Unit*, 1954, No. 217/54.
- LINCOLN, R. S., & ALEXANDER, L. T. Preferred patterns of motor and verbal responses. *J. exp. Psychol.*, 1955, 50, 106-112.
- LINDMAN, H. G., & EDWARDS, W. Supplementary report: Unlearning the gamblers fallacy. *J. exp. Psychol.*, 1961, 62, 630.
- MCGILL, W. J. Multivariate information transmission. *Psychometrika*, 1954, 19, 97-116.
- MCGILL, W. J. Serial effects in auditory threshold experiments. *J. exp. Psychol.*, 1957, 53, 297-303.
- MACGILLIVRAY, M. E., & STONE, C. P. Sug-

- gestions toward and explanation of systematic errors made by albino rats in the multiple light discrimination apparatus. *J. genet. Psychol.*, 1930, 38, 484-489.
- MOWBRAY, G. H. Choice reaction times for skilled responses. *Quart. J. exp. Psychol.*, 1961, 12, 193-202.
- MOWBRAY, G. H., & RHOADES, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, 11, 16-23.
- NEIMARK, E. D. Effect of type of non-reinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, 52, 209-220.
- NEWMAN, E. B. Computational methods useful in analyzing series of binary data. *Amer. J. Psychol.*, 1951, 64, 252-262.
- NEWMAN, E. B., & GERSTMAN, L. J. A new method for analyzing printed English. *J. exp. Psychol.*, 1952, 44, 114-125.
- NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, 57, 105-114.
- OLDFIELD, R. C. Apparent fluctuations of a sensory threshold. *Quart. J. exp. Psychol.*, 1955, 7, 101-115.
- OLDFIELD, R. C. Apparent fluctuations of sensory thresholds: Corrigendum. *Quart. J. exp. Psychol.*, 1956, 8.
- OVERALL, J. E., & BROWN, W. L. Recency, frequency and probability in response prediction. *Psychol. Rev.*, 1957, 64, 314-323.
- PRESTON, M. G. Contrast effect and the psychophysical functions. *Amer. J. Psychol.*, 1936, 48, 625-631. (a)
- PRESTON, M. G. Contrast effects and psychophysical judgment. *Amer. J. Psychol.*, 1936, 48, 389-402. (b)
- PRESTON, M. G. The effect of interpolated time intervals upon the contrast effect. *J. exp. Psychol.*, 1936, 19, 706-715. (c)
- REICHENBACH, H. *The theory of probability: An enquiry into the logical and mathematical foundations of the calculus of probability*. Berkeley: Univer. California Press, 1949.
- ROSS, B. M. Randomization of a binary series. *Amer. J. Psychol.*, 1955, 68, 136-138.
- SENDERS, V. L. Further analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1953, 66, 358-374.
- SENDERS, V. L., & COHEN, J. Effect of sequential dependencies on instrument-reading performance. *J. exp. Psychol.*, 1955, 50, 66-74.
- SENDERS, V. L., & SOWARDS, A. Analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1952, 65, 358-374.
- SKINNER, B. F. The processes involved in the repeated guessing of alternatives. *J. exp. Psychol.*, 1942, 30, 495-503.
- SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, 45, 1-40.
- SOLOMON, R. L. A note on the alternation of guesses. *J. exp. Psychol.*, 1949, 39, 322-326.
- TOLMAN, E. C. Purpose and cognition: The determinant of animal learning. *Psychol. Rev.*, 1925, 32, 285-297.
- TUNE, G. S. The non-independence of responses to a two-choice situation. *J. gen. Psychol.*, 1963, in press.
- TURNER, W. D. Intra-serial effects of lifted weights. *Amer. J. Psychol.*, 1931, 43, 1-12.
- VAN DER MEER, H. C. The influence of instructions in a two-choice probabilistic learning task under partial reinforcement. *Acta psychol., Amsterdam*, 1960, 17, 357-376.
- VERPLANCK, W. S., & BLOUGH, D. S. Randomized stimuli and the non-independence of successive responses at the visual threshold. *J. gen. Psychol.*, 1958, 59, 262-272.
- VERPLANCK, W. S., COLLIER, G. H., & COTTON, J. W. Non-independence of successive responses in measurements of visual thresholds. *J. exp. Psychol.*, 1952, 44, 273-282.
- VERPLANCK, W. S., COTTON, J. W., & COLLIER, G. H. Previous training as a determinant of response dependency at the thresholds. *J. exp. Psychol.*, 1953, 46, 10-14.
- WEISS, B., COLEMAN, P. D., & GREEN, R. F. A stochastic model for time-ordered dependencies in continuous scale repetition judgments. *J. exp. Psychol.*, 1955, 50, 237-244.
- WELFORD, A. T. The psychologist's problem in measuring fatigue. In W. F. Floyd & A. T. Welford (Eds.), *Symposium on fatigue*. London, England: Lewis, 1953. Pp. 91-97.
- WELFORD, A. T. Age changes in the times taken by choice discrimination and the control of movement. *Gerontologia*, 1961, 5, 129-145.
- WERTHEIMER, M. An investigation of "randomness" of threshold measurements. *J. exp. Psychol.*, 1953, 45, 294-303.
- WERTHEIMER, M. Can a trained subject judge his auditory sensitivity? *Quart. J. exp. Psychol.*, 1954, 6, 21-22.
- WERTHEIMER, M. The variability of auditory and visual absolute thresholds "in time." *J. gen. Psychol.*, 1955, 52, 111-147.
- YOSHIOKA, J. G. An alternation habit in rats in a simple maze. *J. genet. Psychol.*, 1929, 36, 257-266.
- ZEAMAN, D., & HOUSE, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, 41, 117-186.

## A NOTE ON "ON SIMPLE METHODS OF SCORING TRACKING ERROR"

NICHOLAS J. CARRIERO

*Systems Support Department, Bell Aerosystems Company, Buffalo, New York*

The article deals with the omission of mean and variability data in Poulton's (1962) review. Implications of the error are discussed and the missing data are supplied.

Poulton (1962) has performed a most needed service in his review of simple methods of scoring tracking error, especially for those laboratories that lack computer facilities for analyzing such data. There is, however, one oversight in the report.

Data presented in Table 1, on page 324, for errors in position at reversals, is broken down into right-left and overshoot-undershoot components. While different values are given for the mean *CE* and the *SE* of the *CE* for each of the components, only one value is shown for the mean *SD* and the *SE* of the *SD*. This relationship suggests that the operations, by which the right-left and overshoot-undershoot error components are derived, are such as to result in a constant difference between them with the consequence that both components yield the same *SD*.

A review of Poulton's procedure for determining these errors indicates that while the same absolute values are involved in both errors, there are sign changes from one component to the

other which are not necessarily consistent. Such conditions assure that the  $\Sigma x^2$  in the *SD* formula will be equal for the two measures but do not assure that the  $(\Sigma x)^2$  will necessarily be equal.

These considerations suggested that the previously reported data were either the result of a fortuitous occurrence or that additional data had been inadvertently omitted from the report.

Poulton has informed me that the latter is the case. The *SD*s presented for position error at reversals in the table are for the right-left component. The *SD* values for the overshoot-undershoot component, he reports, are: Preview Version— $M=1.02$ ,  $SE=.10$ ; Slit Version— $M=2.11$ ,  $SE=.27$ .

He also reports that the Slit mean was larger than the Preview mean for each of the 12 subjects ( $p<.001$ ).

### REFERENCE

- POULTON, E. C. On simple methods of scoring tracking error. *Psychol. Bull.*, 1962, **59**, 320-328.

(Received January 29, 1963)

## DETERIORATION AND FACILITATION HYPOTHESES IN SENSORY-DEPRIVATION RESEARCH

CHARLES A. BROWNFIELD

*Mendocino State Hospital, Talmage, California*

2 opposing hypotheses have emerged from research in sensory deprivation: it impairs mental functioning and it facilitates mental functioning. Much evidence exists for either or both cases, but review of some of the literature suggests that whatever the requirements of the human organism for varied external stimulation, mere reduction or increased patterning of input will not alone produce major disruptive or enhancing effects. The multiplicity of reported results make it appear unwise to conclude that sensory or perceptual isolation, as it is currently conceived, results in deterioration or facilitation of mental function until more research is accomplished. Evidence from learning studies, in which the disparate hypotheses are most pronounced, stimulated the present attempt to integrate empirically the phenomena reported in a wide range of studies in sensory deprivation.

The hypothesis that an organism requires not only stimulation but a continually varied sensory input for the maintenance of normal, intelligent, adaptive behavior (Heron, Bexton, & Hebb, 1953) has been subjected to experimental scrutiny by a number of the original investigators since the publication and amplification of the initial studies conducted at McGill University (Bexton, Heron, & Scott, 1954; Doane, Mahatoo, Heron, & Scott, 1959; Hebb, 1955; Heron, 1957; Heron, Doane, & Scott, 1956; Scott, Bexton, Heron, & Doane, 1959). The findings of these investigators seem to have confirmed the hypothesis of cognitive and perceptual deterioration resulting from exposure to a prolonged period of reduced and decreased variation in the total-sensory environment, a condition concomitant with social and perceptual isolation.

Two studies in this context, both conducted at Princeton University (Vernon & Hoffman, 1956; Vernon & McGill, 1957), have concerned themselves with learning in human beings as a measure of intellectual, cognitive and perceptual efficiency, and the hypotheses in both of these were opposed to what might have been expected in view of the McGill work. These predicted that "sensory

deprivation," a misnomer applied to the original isolation procedure, would facilitate the learning process because of the absence of extraneous inhibiting stimuli interfering with the acquisition of new material. The first Princeton study (Vernon & Hoffman, 1956), using a 48-hour isolation period and four subjects, revealed a significantly better performance for the experimental group, a finding consistent with the facilitation hypothesis, in the expected direction. But the second Princeton study (Vernon & McGill, 1957), utilizing a 72-hour period and a larger *N* of 10, not only failed to confirm the findings of the first experiment at the end of the longer period, but also failed to demonstrate a significantly better performance for the isolates after 48 hours.

A third and more recent study at the University of Miami (Arnhoff, Leon, & Brownfield, 1962), utilizing the same learning task as the two Princeton investigations but replicating the sensory and perceptual conditions of the McGill studies, failed to find significant facilitation or decrement in the performance of an experimental group with an *N* of 12. Neither deterioration nor facilitation hypotheses were supported and the investigators concluded that whatever the

requirements of the adult human organism for external and varied stimulation, reduction or patterning of input will not alone produce major disruptive psychological effects. Such results are the product of a complex interaction of personality, anxiety, expectation, situational structuring, and amount and patterning of external sensory input.

The phenomenon of facilitation by reduction of variable sensory input has thus been shown to be untenable under the conditions imposed by the experimental procedures, although some qualitative differences in subjects' performance were still observed. In general, there was less "disagreement" with the McGill findings in the Miami and second Princeton studies than in the first Princeton study, though agreement does not quite describe the results either.

A possible explanation of the observed discrepancies does not appear so much a matter of depth or duration of sensory deprivation as it is of rejecting any suggested facilitation or deterioration of learning simply by reducing varied sensory or perceptual stimulation. The first Princeton study did not attempt to replicate the original McGill conditions, nor was the second essentially the same as the first. Had a replication been approximated at the beginning, the results of the first learning study might have been more congruent with those of the second, and both could have been interpreted as a refutation of either the deterioration hypothesis or the hypothesis of facilitation, as was the case in the Miami study. This outcome is strongly suggested by examination of the experimental conditions of the two Princeton learning studies, as compared with those at McGill where, incidentally, learning was not the primary process scrutinized. The degree of sensory isolation was more severe in the second Princeton experiment and more closely approximated that of the McGill cubicle; the differences, however, may be crucial

ones for understanding the divergent results.

In the second Princeton study, the duration of time equaled that of McGill's. The experimental cubicle contained only a bed, but the first design used both a bed and a chair. The subjects in the first study had free access to both objects; therefore, the degree of movement restriction was liberal when compared with the second study and the McGill experiment where subjects were restricted to lying on the bed only.

A basic, and perhaps critical, difference was that in the two independent series of studies, one group of subjects spent the major portion of their time in total darkness, while the others always had a moderate amount of light-stimulation to the retina. In the second Princeton study, adjective lists were presented by tape recorder in the cubicle antechamber while the subjects sat blindfolded on the edge of the bed; previously, they had been let out of their cubicle for meals, toileting, and testing in the antechamber by the light of a 15-watt red bulb. By contrast, there was always diffuse illumination from a shielded 40-watt white bulb in the McGill study, but translucent goggles prevented patterned vision.

The McGill cubicle was only sound-deadened, whereas the Princeton cubicle was soundproofed. There was constant noise from the ventilating equipment in the first situation, but great pains were taken to produce a considerable sound-loss (80 decibels) in the latter; here again is a condition which differed in such a way as to lead to some speculation about the necessity of having just minimal stimulation to insure organized thinking.

#### OTHER RESEARCH SUGGESTIONS

Doane et al. (1959) provided some important clues for understanding the phenomenon of the occurrence of hallucinations which may, for the purposes

of the present discussion, be regarded as a rough indicator of the severity of disorganized thought processes, as well as the efficacy of the isolation procedure. They had some subjects wear both opaque and translucent masks; two cubicle subjects wore the opaque masks until an hour before the end of the experimental period, when they were then fitted with the translucent mask worn by the others. Of the 11 subjects who wore the translucent mask continuously, 8 developed hallucinations. When the two subjects, only one of whom had reported having weak hallucinations under the opaque condition, were given translucent masks, both had immediate and vividly strong hallucinations. Five subjects, who were among the most persistent hallucinators with translucent masks, were also put into complete darkness. They all reported an immediate increase in vividness of hallucinations, but within 2 hours there was a reported decrease, three having no more hallucinations and two having them greatly decreased. When they were re-exposed to diffuse light, all of them reported that hallucinatory activity returned to the original level of intensity. It was further reported that two of four ambulatory subjects, who were not confined to the cubicle, but who did wear translucent masks, had hallucinations.

The McGill experimenters suggested not only that diffuse light was only one factor in the hallucinatory phenomena, but that a disturbance in function may be greatest in the area in which restriction occurs, and there might also be some spread of effect to other senses, that is, auditory hallucinations may result from reduction of visual stimulation, etc. In other words, there may be some interaction between sense modalities which depend upon external stimulation. If there is any validity in their observations, then under conditions of almost no stimulation, there should be less disorganization, as gauged in this

case by the occurrence and strength of hallucinations, than under conditions of invariant diffuse stimulation. Such an inference is not inconsistent with the outcome of the Princeton learning studies discussed here, since there was no reliably significant difference between the groups in either situation at the conclusion of the experiment. Such was also the case in the Miami study in which, under the monotonous, nearly static conditions approximating the McGill study, no instance of hallucinations, delusions, or other unusual phenomena was reported.

The results of the McGill studies and the two Princeton investigations raise some question about that part of the deterioration hypothesis which suggests only that some stimulation is necessary for the maintenance of normal, intelligent, adaptive behavior, since lack of stimulation in one case resulted in the reduction or disappearance of hallucinations, and in the other no significant differences between experimental and control subjects could be consistently demonstrated. In the sense that no stimulation elicits no response, the McGill hypothesis is correct; this is not intelligent, adaptive behavior, or more precisely, it is not behavior at all. But if no behavior is manifested, then no inferences of disorganization are warranted, since the organism becomes almost "inert" when nothing happens in the environment for a long enough period. It seems, rather, that a minimal, diffuse, monotonous stimulation is one of the primary requisites for disorganization of cognitive and perceptual function; lack of such stimulation by the severest deprivation of input possible may produce nothing but a lack of reactivity, or a kind of psychological "suspended animation." However, the fact that monotony may be a necessary, but not a sufficient condition for disorganization, is demonstrated by the Miami group.

What seems really to be necessary for the maintenance of integrated behavior is varied sensory input. When hallucinations reportedly occur, or at least when they are of maximum vividness, some visible input is always present from one or more sensory sources. Hallucinatory activity may only be an attempt by the particular individual to differentiate and to structure the input so that it has content and meaningfulness, even if the resulting organization is bizarre because of reliance on internal frames of reference in the absence of external ones.

Silverman, Cohen, Bressler, and Shmavonian (1958) described a situation during a sensory-isolation experiment of short duration in which the subject, a female psychology student, was in a totally dark cubicle. Inadvertently, the black paper covering the observation window was moved so that a crack of light became momentarily visible to the girl. She immediately reported that she had perceived what she thought was a Rorschach card and began elaborating an intricate delusory system around this accidental, fleeting stimulus. It was pointed out that this normally negligible incident was sufficient to produce a complex organization of associations about a topic with which the subject was "ego involved." If one isolated stimulus against the background of monotony, or no stimulation, can evoke such complex behavior, then a continually varied sensory environment would undoubtedly seem necessary for the maintenance of such behavior. When variation is lacking, but a constant input is present, in spite of adaptation the organism "behaves" a little differently than if there is relatively no stimulation whatsoever; relatively little, if anything, happens.

Studies in which severe reduction of sensory input is achieved by immersion in water heated to body temperature (Lilly, 1956; Lilly & Shurley, 1958) have demonstrated that hallucinations and disorganization of thinking ensue

with such rapidity that, by the end of 2 or 3 hours, the subject finds it difficult to continue. This is reminiscent of the increased activity reported by the McGill subjects who were placed in darkness after being in diffuse stimulation; the hallucinations began to disappear with the passage of a time period roughly equivalent to that of the water-tank experiments. In both cases, the rapid change from a condition of stimulation to one of relatively no stimulation resulted in increased hallucinatory activity, that is, disorganization of thinking. The change, in itself, may represent a variability of stimulation to which the organism responds with greater effort and energy by attempting to structure perception; when, finally, there is nothing to perceive, the situation becomes threatening. The process of deafferentiation may be so overwhelming that it becomes a threat to one's very experience of existence, a state which, in the final analysis, may only be determined by the ability to respond to stimuli.

Other supporting evidence for change and variation being essential for organized functioning is the light-deprived subjects' responses upon being re-exposed to light (Doane et al., 1959); hallucinatory activity immediately reappeared with the original vividness. Again, this may be a kind of homeostatic attempt to structure perception in the presence of stimulation, but, falling short of its mark because of the lack of anything to differentiate. Time distortions, too, have been reported, so that the tendency appears, generally, to be an underestimation of the duration of isolation (Lilly, 1956; Lilly & Shurley, 1958; Silverman et al., 1958), an expected result if no progression of temporal events is perceived.

It seems crucial here to re-emphasize the fact that the original McGill studies were primarily concerned with the motivational, cognitive, and perceptual changes resulting from monotonous and

boring stimulation, not with actual deprivation or reduction of stimulation; in this sense, they were imposing a patterning of an extreme sort. No one to date has ever succeeded in depriving a human being entirely of visual, auditory, tactual, kinesthetic, and olfactory stimulation simultaneously (to mention only some of the variables which are capable of being externally controlled); all of the studies cited in the literature have only "monotonized" and reduced the intensity of stimuli or increased the uncertainty of stimulus situations, but have not eliminated them completely. External stimulation was always present in the McGill cubicle in the form of low illumination, constant noise from the ventilating system, and no effective precautions against gross movement. While normal levels of sensory, perceptual, and social stimulation were severely reduced, it does not seem appropriate to call this "deprivation," that is, no stimulation whatsoever. Neither does it appear to be a case of "reduced" patterning of stimuli, but more likely "increased" patterning.

What the water immersion experiments (Lilly, 1956; Lilly & Shurley, 1958) did accomplish appears to be a relatively greater degree of stimulus reduction and decreased variation, especially exteroceptively and proprioceptively, not an absolute reduction or elimination, as implied. In the respirator experiments (Mendelson & Foley, 1956) too, stimulation was not absent, but the variety and intensity were markedly reduced and made so monotonous that the situation was not unlike that of the McGill cubicle or Lilly's water tanks; no actual deprivation seems to be involved, except in degree.

In view of all the methods thus far reported to produce sensory, or perceptual isolation (or invariant stimulation in the deprivation sense), it would appear that only two basic procedures are available to researchers in this area.

The first consists primarily of decreasing the variation of stimulus input to a near-static level so that the effect is that of imposing a monotonized structure on the environmental stimulus complex; this is what has been done in most of the experimental studies reported. The second, and perhaps more difficult method, is to deprive the subject of stimulation in one or more sense modalities so that while neurophysiological receptor systems remain intact, they become nonoperational because of the lack of stimulation. Since no stimulation elicits no response, there is reason to suspect that these two different procedures result in qualitatively different outcomes. Attempts to produce this kind of a situation have been partially successful, since only relative reduction of stimulus intensity has been achieved without the complicating side effects of drug anesthesia or amputation. Somewhat more success attends deprivation of visual and auditory input because of the ease with which light and sound stimulation may be experimentally diminished with appropriate conditions and equipment (Doane, 1955; Hebb, Heath, & Stuart, 1954; Nissen, Chow, & Semmes, 1927; Silverman et al., 1958).

#### NEUROPSYCHOLOGICAL CONSIDERATIONS

Brief mention should here be made of some of the research related to the neurophysiological correlates of observed deprivation phenomena. Recent evidence has suggested that the natural response of the brain-stem reticular system to repeated or monotonous sensory stimulation is a rapid adaptation and consequent extinguishing of arousal and attention-directing functions until a novel stimulus is introduced which represents a change in intensity, frequency, or shift to another sense modality (Jasper, 1958). If nonspecific influences are involved in this effect, they may provide a means by which the brain is able to exclude irrelevant afferent information

during the focus of attention (Hernandez-Peon, Scherrer, & Jouvett, 1956). Magoun (1958) has, in fact, shown that conceptions of brain organization have been enlarged by identification of non-specific neural mechanisms lying between sensory and motor systems, and that reciprocal ascending and descending connections between those lying in the brain stem and wide areas of the hemispheres are involved in arousal to wakefulness and alerting to attention. Moruzzi and Magoun (1949) and French (1957) have identified an area in the brain-stem reticular formation which responds in the same nonspecific way to all afferent stimuli from different receptor sources. This response is simply to arouse the brain, not to relay any specific message; its signals are projected to the entire cortex rather than to any one sensory center. This reticular activating system, or RAS, as the investigators called it, is selective to specific stimuli, that is, it makes differential responses to sudden or even continual sensory input, as if it had the ability to discriminate among stimuli. If the RAS does not function normally, as when there is actual tissue damage, a state of consciousness becomes impossible; the intact system, however, can maintain a wakeful state, even in the absence of a cortex. Here, "wakefulness" is not synonymous with "consciousness," since it is possible to be awake, yet conscious of nothing, as is a newborn infant, or even a subject in an isolation cubicle where there may be almost nothing to perceive or differentiate.

In other words, while a stimulus evokes or guides a specific bit of behavior, it also serves the nonspecific function of maintaining a normal state of arousal through the RAS. In a condition of stimulus impoverishment, decreased variation, or, if possible, in more extreme reduction, the sensory impulses, which normally traverse and activate the reticular system, are, operationally, nonexistent (or markedly reduced), thus

producing an effect akin to that which is achieved by a lesion of the brain stem, that is, deafferentation. Since an actual lesion would naturally make it impossible to study the consequent effects on perceptual and cognitive functioning, the method of experimentally controlling the approximate input of external stimulation by the isolation procedure makes it possible to study the effects on intact subjects. Besides, as Bexton et al. (1954) so poignantly observed, college students, the usual population studied in this context, are understandably reluctant to have their brain stems cut.

Research conducted at Duke University Medical Center was concerned with some rather specific neurophysiological reactions of human beings in low-sensory-input environments and have yielded much valuable information about resting catecholamine and hormonal secretion rates, central nervous system activity, peripheral vascular activity, respiration, liminal discriminations, and other functions (Cohen, Silverman, & Shmavonian, 1960a, 1960b). Using Witkins' Rod-and-Frame procedure and Machover's Draw-a-Person Test to discriminate two distinct groups, Cohen, Silverman, Bressler, and Shmavonian, (1958) have also demonstrated significantly different responses from subjects classified as "field dependent" and "body oriented" (field independent); these differences appear to be related to characteristic modes of interacting with the external environment by reliance upon external or internal cues for a frame of reference. It was suggested that individuals who relied on field cues react differently than body-oriented subjects in a situation in which field cues are minimized or lacking. Field-dependent subjects consistently displayed less ability in discriminating sensory cues, remained more centrally aroused, tended to move around more, appeared uncomfortable, struggled more with feelings and fantasies, or denied them, were more

suspicious, projected internal percepts more frequently, and, when interview responses were grouped to get a rough index of "ego function," showed a greater degree of disorganization.

There was also some suggestion that the individual differences reflected in body-field perceptual relationships may not only be related to responses in an environment that exaggerates these differences and leads to more arousal in one group, but the results may also reflect differences in neurohumoral (noradrenaline) and physiological variables which may be either a function of body-field differences or a parallel expression of some underlying central nervous system difference between the two groups (Cohen et al., 1960b).

#### EARLY EXPERIENCE

The literature abounds with reports of experimental investigations designed to study the deleterious effects of early environmental conditions on later behavior (King, 1958). Most of this work has been done with animals, mainly because of the obvious advantages involved in studying organisms whose rate of maturational development is relatively faster than that of human beings. Attempts at human studies are, for the most part, necessarily restricted to observational methods, case histories, and other types of *ex post facto* designs because of the difficulties inherent in controlling conditions of early experience or of manipulating the environment (Davis, 1940; Dennis & Dennis, 1940; Hill & Robinson, 1929; Melzak, 1954; Spitz, 1954, and others). Generally, however, most experimental efforts with animals or observational methods with humans involve the impoverishment of stimulation at various developmental stages in order to assess the later differential effects of such treatment. It is important to note the emphasis placed

upon the imposition of the deprivation conditions at a stage in the organism's development when it has not fully matured. On the other hand, sensory-deprivation research is concerned primarily with the fully matured adult's current response to stimulus reduction, decreased variation, or absence, of input. In many respects, the results of early-experience studies and sensory-deprivation research are comparable, that is, the cognitive, perceptual, and motivational changes reported are similar, but not precisely the same.

Animals exposed to varying conditions of deprivation, restriction, or isolation at different chronological ages may manifest entirely different behavioral responses, depending upon the particular developmental stage at which they undergo the treatment (King, 1958). In sensory-deprivation research, the organism's ability, or inability, to adapt to the stresses of the contemporary stimulus situation, or to no stimulation at all, is of primary concern, and reports of the consequences of such stresses are very consistent with one another, depending, of course, upon the type of deprivation employed. Whereas early experience almost always appears to result in permanent, usually deleterious, alteration of what would otherwise have been typical behavior for the particular species under consideration, sensory or perceptual impoverishment, or invariability, seems to have its greatest effect upon the mature organism at the time it is being subjected to these conditions; perservation of effects vary with the intensity and duration of the isolation, as well as with the individual subject but no permanent or lasting changes have ever been reliably reported. This is not to say that, given specific stimulus conditions in sufficient amounts for undetermined durations and with appropriate subjects, lasting deteriorative or facilitative changes are not found; but it may very well be

that, once we succeed in establishing a reliable body of knowledge about the phenomena and learn to control them, permanent changes may be experimentally accomplished. So far as is now known, perseveration of effects seem to last, at most, 6 weeks (Bexton, Heron, & Scott, 1954), but, if at all, more likely a matter of hours or days (Arnheoff et al., 1962; Cohen et al., 1958; Doane et al., 1959; Heron et al., 1956; Lilly & Shurley, 1958; Mendelson & Foley, 1956; Vernon et al., 1956, 1957).

A notable exception to the idea of cognitive or perceptual change being deleterious and of short duration after isolation is the claim of increased receptivity to the benefits of psychotherapy and concomitant improvement of psychotic patients who were exposed to conditions approximating the McGill study (Azima & Cramer, 1956; Harris, 1959). The efficacy of the isolation procedure in this case, and its relationship to the improvement or facilitation of recovery from a psychotic state may be questionable for the obvious reason that no conclusive evidence that any change observed was the direct consequence of the isolation. Practical experience in the clinical setting, however, has frequently led to the use of restraint or isolation of some types of agitated patients. The results of restraint or isolation would suggest that such procedures tend to reduce confusing internal perceptive and cognitive activity, bringing about a state of relative calm, with increased responsiveness to the ensuing stimulation of psychotherapy. It is interesting to speculate that a disturbed individual may very well become almost "normal" in his response to isolation by becoming more tractable and receptive to measures which are known to be effective with more normal people. Until we know more about sensory and perceptual isolation, however, such a notion must remain within the realm of speculation only.

#### INDIVIDUAL DIFFERENCES

That individual responses may differ according to personality differences has been amply demonstrated by the fact that some subjects in isolation show one kind of response, for example, hallucinations, disorganization, performance deficits, and others show an entirely different kind of response, for example, no hallucinations, little, if any, disorganization, and, perhaps, some facilitation of performance. Not all the McGill subjects reacted to decreased variation with discomfort or annoyance, nor did Cohen et al. (1960a, 1960b) report similar uniform reactions with their subjects at Duke. Twenty-nine people began the McGill study, while only 18 were sufficiently "comfortable" enough to finish; there must have been both personality and motivational differences between those who did and those who did not finish. Similar influences have been drawn in almost all other accounts. Silverman and his associates (1958) reported that, in a pilot study, individuals prediagnosed as schizoid personalities were more content and even liked the situation more than normal subjects; body-oriented subjects were less disturbed than field-dependent ones, and sex-role differences were manifested in projective responses. These same investigators gave anecdotal evidence of specific ego involved behavior; subjects projectively perceived common objects in their studies or employment in a single, accidental, fleeting exposure to a crack of light, for example, a psychology student saw a Rorschach card, a medical student saw an X ray, etc. (Silverman et al., 1958). The work of Azima and Cramer (1956) and Harris (1959) lends credence and some suggestion of validity to these observations in the sense that schizophrenics can be considered as a specific class of personality syndrome in that they display commonly similar types of responses to sensory isolation, as

distinct from the normal subjects thus far used in other studies, though most of the latter studies have tended to use unclassified, unevaluated volunteers or "warm bodies" for subjects.

Interestingly enough, it has been related (Parry, 1960) that Soviet astronauts ("cosmonauts"), after being exposed to isolation and weightlessness as part of their preparatory training for space flight, are said to have enjoyed this condition so much that all of them confessed that they could think of no better way to spend a vacation. It may very well be that schizoid personalities, as reported by the Duke group, and as borne out by the clinical studies mentioned above, might make the best space pilots; if being body oriented and concrete minded are the most suitable prerequisites for survival in space, then such seemingly atypical personalities might be better equipped to do so. Personality differences may account for differential motivation in volunteering for space-flight projects, but it is equally conceivable that motivation can produce the particular personality which finds need fulfillment in such ventures. Such considerations have opened whole new avenues of research, not only concerned with space travel per se, but intimately related to the entire field of personality dynamics and motivation.

Another way in which sensory and perceptual isolation studies are related to the problems of individual differences is the widely accepted notion, very frequently utilized in other kinds of studies, of need or deficit. Just as deprivation of nourishment results in increased activity, which functionally produces coordinated motor responses designed to place the organism in a more favorable situation to acquire nourishment, if it is to survive, stimulus deficit also produces behavior designed to acquire stimulation, if, as Hebb (1955) suggests, it is to sur-

vive. Nutrient qualities of stimulation could be responsible, in large measure, for the need for stimulation, or stimulus-seeking behavior manifested by subjects who hallucinate, report discomfort, or actively attempt to produce situations which will ultimately lead to some form of real or hallucinated sensory-gratification; the perceptual and cognitive changes noted during and after isolation may be by-products or artifacts of deficiency, just as such changes are observed in subjects undergoing semistarvation (Franklin, Schiele, Brozek, & Keys, 1948). It appears quite possible that one can be "starved" for stimulation, as well as for food, and the responses associated with such a state seem at least dynamically similar to other kinds of starvation. Exploratory behavior or curiosity might be thought of as a kind of "snack" with novel stimuli for its substance, just as a fully food-satiated individual might partake of an extra delicacy or two between meals. Not being able to obtain food, or stimulation, because it has been irretrievably withdrawn, it then becomes an extremely important prerequisite for integrated, intelligent, adaptive behavior. Motivation, then, is quite prominent and relevant in the analysis of responses to sensory deprivation, invariant patterning, overstimulation, and monotony.

#### REFERENCES

- ARNHOFF, F. N., LEON, H. V., & BROWNFIELD, C. A. Sensory deprivation: Its effects on human learning. *Science*, 1962, 138, 899-900.
- AZIMA, G., & CRAMER, F. J. Effects of partial isolation in mentally disturbed individuals. *Dis. nerv. Sys.*, 1956, 17, 117-121.
- BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, 8, 70-76.
- COHEN, S. I., SILVERMAN, A. J., BRESSLER, A. J., & SHMAVONTIAN, B. M. Practical and theoretic difficulties in "isolation" studies. Presented at Office of Naval Research symposium on sensory deprivation, Boston, June 1958.

- COHEN, S. I., SILVERMAN, A. J., & SHMAVONIAN, B. M. Psychophysiological mechanisms of stress responsivity. Semiannual report on Air Force scientific research, April 1959–April 1960, Duke University Medical Center, AF 49 (638)–354. (a)
- COHEN, S. I., SILVERMAN, A. J., & SHMAVONIAN, B. M. Psychophysiological mechanisms of stress responsivity. Semiannual report on Air Force scientific research, April 1960–October 1960, Duke University Medical Center, AF 49 (638)–354. (b)
- DAVIS, K. Extreme social isolation of a child. *Amer. J. Sociol.*, 1940, 1, 554–565.
- DENNIS, W., & DENNIS, M. G. The effect of cradling practices upon the onset of walking in Hopi children. *J. genet. Psychol.*, 1940, 56, 77–86.
- DOANE, B. K. Changes in visual function with perceptual isolation. Unpublished doctoral dissertation, McGill University, 1955.
- DOANE, B. K., MAHATOO, W., HERON, W., & SCOTT, T. H. Changes in perceptual function after isolation. *Canad. J. Psychol.*, 1959, 13, 210–219.
- FRANKLIN, J. C., SCHIELE, B. C., BROZEK, J., & KEYES, A. Observation on human behavior in experimental semi-starvation and rehabilitation. *J. clin. Psychol.*, 1948, 4, 28–45.
- FRENCH, J. D. The reticular formation. *Scient. American*, 1957, 196, 54–60.
- HARRIS, A. Sensory deprivation and schizophrenia. *J. ment. Sci.*, 1959, 105, 235–237.
- HEBB, D. O. The mammal and his environment. *Amer. J. Psychiat.*, 1955, 111, 826–831.
- HEBB, D. O., HEATH, E. S., & STUART, E. A. Experimental deafness. *Canad. J. Psychol.*, 1954, 8, 152–156.
- HERNANDEZ-PEON, R., SCHERRER, H., & JOUVET, M. Modification of electrical activity in unanesthetized cats. *Science*, 1956, 123, 331–332.
- HERON, W. The pathology of boredom. *Scient. American*, 1957, 196, 52–56.
- HERON, W., BEXTON, W. H., & HEBB, D. O. Cognitive effects of a decreased variation in the sensory environment. *Amer. Psychologist*, 1953, 8, 366. (Abstract)
- HERON, W., DOANE, B. K., & SCOTT, T. H. Visual disturbance after prolonged perceptual isolation. *Canad. J. Psychol.*, 1956, 10, 13–18.
- HILL, J. C., & ROBINSON, B. A case of retarded mental development associated with restricted movements in infancy. *Brit. J. med. Psychol.*, 1929, 2, 268–277.
- JASPER, H. H. Reticular-cortical systems and theories of the integrative action of the brain. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958.
- KING, J. A. Parameters relevant to determining the effect of early experience upon the adult behavior of animals. *Psychol. Bull.*, 1958, 55, 46–58.
- LILLY, J. C. Mental effects of physical restraint and of the reduction of ordinary levels of physical stimuli on intact, healthy person. *Psychiat. res. Rep.*, 1956 (June), No. 5.
- LILLY, J. C., & SHURLEY, J. T. Experiments in solitude in maximum achievable physical isolation with water suspension of intact, healthy person. Paper read, in part, Symposium on sensory deprivation, Harvard Medical School, Boston, June 1958.
- MAGOUN, H. W. Non-specific brain mechanisms. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: Univer. Wisconsin Press, 1958.
- MELZAK, R. The genesis of emotional behavior. *J. comp. physiol. Psychol.*, 1954, 47, 166–168.
- MENDELSON, J. H., & FOLEY, J. M. An abnormality of mental function affecting patients in a tank-type respirator. *Trans. Amer. Neurol. Ass.*, 1956, 81, 134–138.
- MORUZZI, G., & MAGOUN, H. W. Brain-stem reticular formation and activation of the EEG. *EEG clin. Neurophysiol.*, 1949, 1, 455–473.
- NISSEN, H. W., CHOW, K., & SEMMES, J. The effects of restricted opportunity for tactual, kinesthetic and manipulative experimentation on behavior of a chimpanzee. *Amer. J. Psychol.*, 1927, 39, 283–296.
- PARRY, A. Here come the cosmonauts: Behind the scenes of Russia's man into space program. *Sci. Dig.*, 1960, 48, 23–28.
- SCOTT, T. H., BEXTON, W. H., HERON, W., & DOANE, B. K. Cognitive effects of perceptual isolation. *Canad. J. Psychol.*, 1959, 13, 200–209.
- SILVERMAN, A. J., COHEN, S. I., BRESSLER, B., & SHMAVONIAN, B. M. Hallucinations in sensory deprivation. Presented at Symposium on hallucinations, American Association for the Advancement of Science, December 1958, Washington, D. C.
- SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child*, 1954, 1, 53–74.
- VERNON, J., & HOFFMAN, J. Effects of sensory deprivation on learning rate in human beings. *Science*, 1956, 123, 1074.
- VERNON, J., & MCGILL, T. E. The effects of sensory deprivation upon rote learning. *Amer. J. Psychol.*, 1957, 70, 637–639.

(Received February 25, 1963)

# THREE THEORETICAL INTERPRETATIONS OF ORDER EFFECTS IN PERSUASIVE COMMUNICATIONS<sup>1</sup>

ROBERT E. LANA

*Alfred University*

The "set," "reinforcement," and "sensory variation" hypotheses are reviewed in terms of their ability to predict the results of order effects in the persuasive situation where opposed arguments on the same topic are utilized. The set hypothesis is most successful when the communicative materials presented are unfamiliar to the S. The sensory variation hypothesis is most successful when the topics of the communications are concerned with familiar social issues.

Although interest in the position of opposed communications and their effectiveness in changing opinions was present at least as early as 1925 (Lund, 1925), research activity was begun in earnest only recently. Carl Hovland and his associates (1957) produced a volume of research and theory which probably stirred other researchers to re-examine this problem area. A body of research on order effects as a result of persuasive communications has since developed. It is the purpose of this paper to review three major theories which attempt to explain the empirical results. Primacy refers to the success in changing opinion of the initial argument of two opposed communications. Recency refers to a similar success of the argument presented second.

One of the initial explanations of the influence of primacy when opposed communications are successively presented to a subject is that suggested by Luchins (1958) and Anderson and Barrios (1961), having its immediate history in an experiment performed by Asch in 1946. The concept utilized was that of "set" or *Einstellung*, much in the same sense in which the terms were used by

nineteenth century psychologists. The fundamental idea is that a subject, confronted with unfamiliar material, will establish a tendency to react later in terms of the initial material rather than in terms of different material presented after it. Under this hypothesis, it must be assumed that no organization exists in the subject with respect to the communication presented to him. The initially unfamiliar material supplies such an organization which influences his subsequent responses (opinion expression via questionnaire). Thus, presentation of some communication, which supports a given opinion or establishes a particular description, will render succeeding material relatively ineffective in changing opinion compared with the first presentation. This effect should be maximal the greater the temporal contiguity of the two presentations since forgetting, and interference by external factors, would be minimal. This hypothesis depends heavily upon the condition that the presentation of a topic with which the subject is unfamiliar be truly established, since any previous contact by the subject with the material would constitute violation of this condition. It thus becomes necessary in studies of opinion change to devise topical material which is as unique as possible for the subject. Current popular social issues are, there-

<sup>1</sup>This study was supported by Research Grant MH 6926-01 from the National Institute of Mental Health, United States Public Health Service.

fore, not suitable for use as topics of communications. The first elements of the communication regarding some social issue may simply aid the recall of already-formed opinions about the matter, and thus the set interpretation will not be applicable. The possibility, however, exists that a set may be formed in the given experimental situation in spite of any recall which might be induced in the subject by the initial communication. An examination of the empirical evidence bearing on this point will be made below.

Luchins (1957a, 1957b) presented two different descriptions of the same individual to high school and college students. These descriptions were characterized by being polar in nature on an introversion-extraversion continuum. Hence, they were essentially opposed communications. Complete lack of familiarity with the subject matter was assured since the individual who was being described ("Jim") was fictitious. The subjects were then asked to describe Jim in their own words and answer questions about his personality characteristics. The presentation of the opposed descriptions was juxtaposed for the various experimental and control groups. Luchins found consistent primacy effects. He favored a set explanation of his results, in that the initial description directed the later opinions of the subjects in much the same way that a particular initial solution to a problem affects attempts to solve later similar problems.

The effects on opinion change of these initial impressions may be minimized by interposing, throughout the experimental process, directions which change the set. For example, Luchins (1957a, 1957b) interposed comments such as, "I want each of you [subjects] to suspend judgment of the individual ["Jim"] about whom you are to read until you have

completely finished reading all that is written about him . . . ." With comments such as these, primacy effects were reduced in magnitude and, depending upon how late in the procedure the comment or text was introduced, recency effects were sometimes obtained. Luchins (1958) also found a recency effect when a questionnaire was administered after each of two opposed communications had been delivered and the topic included different descriptions of the same individual. These results were opposite to the results of Lund (1925) who found primacy effects under the same test conditions. However, Luchins' communications were markedly different from Lund's. Hovland and Mandell (1957) could not replicate Lund's results. The effects of differences in the communicative materials will be discussed below. Thus, Luchins' results indicate consistent primacy effects which would be predicted by a set theory.

Anderson and Barrios (1961) have also reported experiments which tend to support the set hypothesis. They examined order effects over a sequence of communications on separate issues. Personality adjectives were used as the communications. Both Asch (1946) and Luchins (1958) predict primacy effects for initial trials and recency effects for later trials. The subjects were instructed to attempt to form an impression of an imaginary person who was being described by a series of six adjectives, all relating to his personality. Favorable and unfavorable adjectives were presented in various sequences to the experimental groups and significant primacy effects were observed in all groups. That is, the adjectives presented first were more influential than those presented later. Also, there was a significant decline in primacy effect over trials of presentation of various adjective sets. The authors concluded that when

only two sets of adjectives (communications) are used (as in other experiments and their own Experiment II) the true order effect is obscured. Hence, several communications are necessary for a primacy effect to develop. Primacy, however, decreases over communication sets. This decrease in primacy is explained by the authors either in terms of a progressive loss in interest on the part of the subject in the task, or an adaptation process whereby previous material is integrated by the subject with practice; or it may be that with increasing exposure a set develops to take increasing cognizance of all the words in the set, thus destroying any order effect that might have been originally present in the first few trials. There are several experiments where the set hypothesis does not predict the results obtained. This may be because the communicative situation in such experiments does not contain the essential conditions necessary for correct predictions from a set hypothesis.

Studies have appeared in the literature which utilize a different procedure from that of Luchins and Anderson and Barrios. In the studies by Hovland and Mandell (1957), Lana (1961), and Thomas, Webb, and Tweedie (1961), the communications consisted of prose passages on topics of greater or lesser importance involving actual social issues (atomic submarines, cancer research, nuclear weapons, Picasso, etc.). Pretests and posttests, usually consisting of identical questionnaires, were also used. These studies did not present the subjects with initial set-producing communications as in the Luchins and Anderson and Barrios studies since the communications themselves pertained to topics with which, it is safe to assume, the subjects had had some previous contact. Also, the fact of the pretest would serve, by its content, as a device aiding recall of information about the topic.

In a series of studies by Lana (1961, 1962) and his associates a consistent primacy effect was found when familiarity with a topic was high and a recency effect was evident when familiarity was low.

In one study (Lana, 1961) familiarity was manipulated by differentially providing information about an unfamiliar topic to two groups, while a third group received no such information. The "very familiar" group yielded a primacy effect and the "unfamiliar" group a recency effect. The middle "familiar" group yielded no significant order effect. Rosnow (1962) successfully replicated Lana's results. In this instance existing familiarity with the topic was measured by a pretest.

Thomas, Webb, and Tweedie (1961) performed three experiments to examine the effects of familiarity with a controversial issue on opinion change following two opposed successive arguments. The authors argued that a primacy effect can only be expected to appear when the topic of communication is first exposed to the subject in the experimental situation. Subjects were made familiar with certain facts associated with the communications on the use of cancer serum while another group was not familiarized with this topic. The usual orders of presentation were used for the various groups. There was no difference between the familiar and the unfamiliar groups with respect to primacy. The second and third experiments utilized delayed posttests. There was no significant primacy effect in either the familiar or the unfamiliar groups. Although not specifically reported, presumably no significant recency effects were present in any of the studies. Their results confirm ours in the experiments mentioned above on familiarity. There is no primacy effect evident with unfamiliar groups. Actually, in the studies by Lana

and Rosnow a recency effect is in greater evidence. One of the differences between the Thomas, Webb, and Tweedie studies and the Lana experiment is that the degree of controversiality of the topics is probably not the same in both sets of procedures. A more controversial topic was utilized in the Lana experiment than in the Thomas, Webb, and Tweedie study. This conclusion seems to be warranted by the fact that Lana's subjects (1962), in another context, rated cancer research (used by Thomas, Webb, & Tweedie, 1961) to be of medium controversiality compared to nuclear weapons, Khrushchev, etc. However, Thomas, Webb, and Tweedie showed that the topic of use of cancer serum was perceived to be controversial by the subjects. These results are superficially opposite to those predicted from the set hypothesis. From the set hypothesis it follows that the more familiar an individual is with a given topic the less likely he is to establish a set on the communication presented first which should result in a recency effect. Conversely, the less familiar an individual is with a topic the more likely he is to establish a set with respect to the first communication thus yielding a primacy effect. However, since the topics used by Lana and Thomas, Webb, and Tweedie were already familiar to college subjects, some familiarity already existed even for those groups termed unfamiliar in the various experiments.

Other studies by Lana (1962) and Schultz (1963b) where controversy of and interest in the topic and awareness of the intent of the experimenter by the subject were manipulated, yielded perplexing results. If it may be assumed that "controversy," "interest," and "awareness" may all be related to the formation of set in the communication situation, then this hypothesis is apparently unable to predict the empirical

results. Schultz (1963b) found that groups that had been made strongly aware of the intent of the communicator produced no order effect when confronted with opposed communications. A moderately-aware group yielded a primacy effect, and a recency effect was present in a group of low awareness. According to the set hypothesis, the low-aware group should have yielded the primacy effect and the high-aware group the recency effect. Lana (1962) showed that primacy appeared in groups exposed to a very controversial topic and where there was great interest in the topic. The set hypothesis would have predicted recency effects for these groups or no effect at all. At this point it is perhaps better to conclude that the set hypothesis is not structured to explain order effects when communications concerning familiar social problems are involved, rather than to conclude that it is made tenuous by such studies.

A "reinforcement" or "conditioning" explanation of order effects has been utilized in certain contexts by McGuire (1957) and by Rosnow (1962). The prediction of a particular order effect depends upon the ability of the experimenter to create or locate a rewarding or punishing group of stimuli in the communication situation. It is problematic whether or not reward or punishment as reinforcers can be located as factors within the experiments described above. An individual may be convinced of an argument to either increase or decrease support of, for example, cancer research in the face of the opposite position without a reward or a punishment being contingent on his taking such a stand.

For McGuire (1957) a "conditioning trial" occurs when a subject agrees with the communication (via questionnaire) and this agreement is reinforced. If there is no reinforcement of the agreement, an "extinction trial" has occurred.

It was necessary to establish that the topics had either pleasant or unpleasant contingencies involved with a belief in any of them. McGuire did this by administering appropriate questionnaires. He demonstrated that "the effect of a source's earlier communications on the persuasiveness of his later messages depends in large measure on the extent to which agreement with those earlier communications was rewarding for the recipient." Reward consisted of the situation where results previously judged to be desirable by the subjects (e.g., federal aid to education) were regarded as likely to occur by the communicator. It is to be noted that McGuire's experiment dealt with order effects within a communication rather than with order effects concerning opposed communications as used in the above studies. However, Rosnow's (1962) experiment was concerned with successive opposed arguments. He created a punishing situation, the relief from which he characterized as reinforcement. Anxiety was created among high school students by administering an unannounced quiz in their psychology class. Two days later pro and con arguments concerning civil defense were presented either before or after anxiety about quiz results had been reduced by informing the students that their grades would not count on their records. Throughout the study, opinion questionnaires were administered as well as a mood questionnaire to check the effects of the anxiety-producing situation. Rosnow's principal hypotheses were that reinforcement occurring before the presentation of opposed arguments would yield a primacy effect, and that reinforcement occurring after the presentation of opposed arguments would yield a recency effect. These predictions were made on the supposition that contiguity between the reinforcement and either the first- or second-presented argu-

ment would strengthen the tendency to respond in that direction. The results indicated that, although anxiety had been successfully created, there were no consistent order effects resulting from reinforcement (anxiety reduction) before or after the presentation of the arguments. Since no clear-cut effects were present, no further information concerning a reinforcement interpretation of order effects is available at this time.

The most recent attempt to explain order effects in persuasive communications has been made by Schultz (1963a). Schultz developed four postulates and two corollaries from a physiological "sensory-variation" hypothesis based on the work of Hebb (1955), Scott (1957), Fiske and Maddi (1961), Lindsley (1957), and Malmo (1959). The principal idea is that human beings are disposed to seek high activation and, hence, to react more strongly to novel stimuli than to stimuli with which they are familiar, since novel stimuli provide higher cortical activation, other things being equal. We need not be detained by this point (for details see Schultz, 1963a). From this general point of view Schultz derives the four following hypotheses:

1. When an individual is exposed to a topic for the first time (as in Luchins, Asch, Anderson and Barrios, called "true" primacy studies by Schultz) the first communication is perceived as novel and mediates increased cortical activation. A second (opposed) communication on the same topic provides little activation compared with the first communication and, hence, in order to maintain activity, the individual reacts in accordance with the position advocated by the initial communication. A primacy effect results.

2. If the subject is required to register his opinions after each (pro and con) communication, then, in order to main-

tain the level of activation induced by the first communication, he will respond on the second questionnaire in opposition to his response on the first questionnaire. A recency effect will result.

3. Time or task interpolation between first and second presentations of the opposed communications serve to dissipate activation level induced by the first communication and, therefore, the second communication will be more effective and a recency effect or no effect at all will occur. The second communication has the characteristics of a novel stimulus under these conditions.

4. When studies utilize communications which are not initial (as when they are about current social issues, such as integration, aid to education, etc.; called "experimentally induced primacy studies" by Schultz) then the first communication will not produce as high an activation level as if it were an initial exposure. The first communication may still be somewhat novel, however, and may determine some increased activation and attention. Hence, a less pronounced primacy effect (as compared with true primacy studies) or no effect at all is predicted.

Schultz (1963a) then appends two corollaries (*a* and *b*) to his fourth postulate. According to *a*, if, under the conditions of 4 above, the social issue has an intense stimulus value along dimensions such as familiarity, interest, or controversiality, then the dynamics of the communication situation parallels Postulate 1 above and a more pronounced primacy effect should occur. Schultz makes the implicit assumption that dimensions such as interest in the topic and degree of controversy as perceived by the subject have the same systematic status as familiarity of the subject with the topic, which is the dimension involved in the true primacy studies. Corollary *b* states that in the situation

described in 4a a recency effect should occur when a questionnaire is interpolated between the two communications.

Since Schultz makes the same predictions as set theorists (derived from different initial assumptions) for his true primacy studies his success is equal to theirs.

Difficulty arises with Schultz's predictions 4, 4a, and 4b in studies which are not "initial" in the sense used throughout this paper.

The earliest study cited in this context is the one by Lund in 1925. Schultz (1963a) predicts a recency effect by Postulate 2, but Lund reported a primacy effect with questionnaires presented after each communication (pro and con). Cromwell (1950), using a single questionnaire after the presentation of both arguments, indicated a recency effect was present, but Schultz predicts primacy. Hovland and Mandell (1957) replicated Lund's study using a different topic and their results are also inconsistent with Schultz's predictions. The success of his model with induced primacy studies increases somewhat when applied to some of the results of Lana (1961, 1962). From Schultz's Postulates 4 and 4a, primacy is predicted for groups with high familiarity with the topic and when the topic is very controversial. However, Schultz does not report that, where his Postulates 4 and 4a predict the results for familiarity and controversy, they do not adequately handle Lana's results (1962) in an experiment where high-interest groups yielded no effect and a medium-interest group yielded a primacy effect. Schultz must predict primacy for the high-interest groups and no effects or recency for the medium-interest groups. In another study (Lana, 1962) high controversy, in interaction with method of pretesting, yielded no order effect and medium controversy

yielded a significant primacy effect. Schultz simply cannot predict all of Lana's results as he implies in his paper. However, of the three theories reviewed, it seems clear that the sensory-variation hypothesis comes closest to predicting order effects where the topics represent current social issues.

## REFERENCES

- ANDERSON, N. H., & BARRIOS, A. A. Primacy effects in personality impression formation. *J. abnorm. soc. Psychol.*, 1961, 63, 346-350.
- ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, 41, 258-290.
- CROMWELL, H. The relative effect on audience attitude of the first versus the second argumentative speech of a series. *Speech Monogr.*, 1950, 17, 105-122.
- FISKE, D. W., & MADDI, S. R. *Functions of varied experience*. Homewood, Ill.: Dorsey Press, 1961.
- HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, 62, 243-254.
- HOVLAND, C. I., MANDELL, W., CAMPBELL, ENID H., BROCK, T., LUCHINS, A. S., COHEN, A. R., MCGUIRE, W. J., JANIS, I. L., FEIERABEND, ROSALIND L., & ANDERSON, N. H. (Eds.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957.
- HOVLAND, C. I., & MANDELL, W. Is there a "law of primacy in persuasion"? In C. I. Hovland et al. (Eds.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957. Pp. 13-22.
- LANA, R. E. Familiarity and the order of presentation in persuasion. *J. abnorm. soc. Psychol.*, 1961, 62, 573-577.
- LANA, R. E. Order effects in persuasive communications. Progress report, May 1962, United States Public Health Service, National Institute of Mental Health, Research Grant M-4830.
- LANA, R. E. Controversy of the topic and the order of presentation of persuasive communications. *Psychol. Rep.*, 1963, 12, 163-170. (a)
- LANA, R. E. Interest, media and order effects in persuasive communications. *J. Psychol.*, 1963, 56, 9-13. (b)
- LANA, R. E., & ROSNOW, R. L. Subject awareness and order effects in persuasive communications. *Psychol. Rep.*, 1963, 12, 523-529.
- LINDSLEY, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957.
- LUCHINS, A. S. Experimental attempts to minimize the impact of first impressions. In C. I. Hovland et al. (Eds.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957. Pp. 62-75. (a)
- LUCHINS, A. S. Primary-recency in impression formation. In C. I. Hovland et al. (Eds.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957. Pp. 33-61. (b)
- LUCHINS, A. S. Definitiveness of impression and primacy-recency in communications. *J. soc. Psychol.*, 1958, 48, 275-290.
- LUND, F. H. The psychology of belief: IV. The law of primacy in persuasion. *J. abnorm. soc. Psychol.*, 1925, 20, 183-191.
- MCGUIRE, W. J. Order of presentation as a factor in "conditioning" persuasiveness. In C. I. Hovland et al. (Eds.), *The order of presentation in persuasion*. New Haven: Yale Univer. Press, 1957. Pp. 98-114.
- MALMO, R. B. Activation: A neurophysiological dimension. *Psychol. Rev.*, 1959, 66, 367-386.
- ROSNOW, R. L. Opinion change and order of presentation in experimentally manipulated anxiety and in natural familiarity with a topic. Unpublished doctoral dissertation, American University, 1962.
- SCHULTZ, D. P. Primacy-recency within a sensory variation framework. *Psychol. Rec.*, 1963, 13, 129-139. (a)
- SCHULTZ, D. P. Time, awareness, and order of presentation in opinion change. *J. appl. Psychol.*, 1963, 47, 280-283. (b)
- SCOTT, T. H. Literature review of the intellectual effects of perceptual isolation. Report No. HR66, July 1957, Department of National Defense, Defense Research Board, Canada.
- THOMAS, E. J., WEBB, S., & TWEEDIE, J. Effects of familiarity with a controversial issue on acceptance of successive persuasive communications. *J. abnorm. soc. Psychol.*, 1961, 63, 636-659.

(Received March 26, 1963)

# Psychological Bulletin

## DISCRIMINATION LEARNING SET IN RHESUS MONKEYS<sup>1</sup>

HAYNE W. REESE

*State University of New York at Buffalo*

The rate of acquisition of discrimination learning set in rhesus monkeys is related to amount of within-problem learning, problem difficulty, and amount of separation between stimuli and response loci. Experimentally naive monkeys have a weak tendency to approach novel stimuli, but because of generalization from the positive to the negative stimulus within a problem, experimentally sophisticated monkeys tend to approach familiar negative stimuli rather than new stimuli. Early in training there are more errors following a first-trial error than following a correct first-trial response, but later this trend is reversed. All of the data can be accounted for by Hull-Spence theory, with a small number of assumptions added.

This article has two main purposes: first, to review the research on discrimination learning set in rhesus monkeys, and second, to outline a theoretical analysis of learning set based on Hull-Spence learning theory.

The few studies of learning-set formation that used infrahuman primates other than rhesus monkeys as subjects are excluded because the interpretation of the obtained species differences is complicated by the use of different experimental procedures (see Harlow, 1959). (Data of irradiated monkeys and monkeys with brain lesions have also been excluded.) The review is limited to discrimination learning set (Harlow, 1949) because it has been more thoroughly investigated than the other kinds of learning set, such as oddity learning set (e.g., Meyer & Harlow, 1949; Moon & Harlow, 1955) and reversal learning set (e.g., Harlow, 1949; Meyer, 1951), and in spite of the

use of many different experimental procedures, a "standard" technique for establishing a discrimination learning set (hereafter called "learning set") can be described. In the standard learning-set technique, the subject is given a series of problems that have a common basis of solution; each problem involves: (a) simultaneous discrimination between two stimulus objects, with a different pair of stimuli for each problem; (b) blind baiting of the correct stimulus; (c) reward for every correct response; (d) a non-correctional procedure; and (e) some small, fixed number of trials (on each problem). Improvement from problem to problem in the performance level on a given within-problem trial, usually Trial 2, is used as a measure of inter-problem transfer or learning-set formation.

Current analyses of learning-set data deal with "error factors" (Harlow, 1949, 1950, 1959), "hypotheses" (Levine, 1959), or "strategies" (Bowman, 1961). These are response-defined terms, referring to systematic sequences of responses,

<sup>1</sup> The writer is indebted to several colleagues, especially B. R. Bugelski, John W. Davenport, and Billey Levinson, who read and criticized earlier versions of this paper.

as did Krechevsky's (1932) term "hypothesis." Naming the response sequences does not explain them, however, and no mechanism to account for the development and operation of strategies, hypotheses, or error factors has been suggested, although several theorists have dealt with the problem (e.g., Harlow, 1950, 1959; Harlow & Hicks, 1957; Levine, 1959; Restle, 1958, 1960, 1962).

There have been suggestions in the research literature on learning sets in monkeys that various findings can be accounted for by traditional learning theory. Harlow (1950), for example, noted that Hull's (1943) concept of reactive inhibition could account for one of the error factors ("response-shift") [although he later rejected the notion (Harlow, 1959)], and Riopelle, Cronholm, and Addison (1962) concluded that:

the critical factors in gradual learning-set formation . . . are to be found in cue stimuli and the interproblem relations rather than in the acquisition of supporting habits or the extinction of antagonistic habits [p. 277].

("Supporting habits" may be categorized as performance sets, and "antagonistic habits" as error factors or hypotheses.) Similarly, Brush, Mishkin, and Rosvold (1961) and Harlow (1959) suggested that the approach to or avoidance of particular stimulus objects may be due in part to generalization from other similar stimulus objects that were positive or negative earlier in training.

The notion that an analysis may be based on traditional learning theory is not new (see Bugelski, 1956), but no such analysis has previously been attempted.

#### REVIEW OF THE LITERATURE

The writer has reinterpreted the data of many of the studies, requiring that the experimental designs be presented in more detail than is customary in reviews. Especially in the section on Reactions to

Novel Stimuli, the new interpretations yield conclusions that were not originally apparent.

#### *Experimental Variables*

*Number of Trials per Problem.* According to Levine, Levinson, and Harlow (1959) the rate of learning-set formation is a function of the number of trials presented, regardless of how these trials are organized into problems, provided the number of trials per problem is between 3 and some number greater than 12 but less than 50. When more trials than the upper limit are given on each problem, the rate of learning-set formation is retarded. Levine, Harlow, and Pontrelli (1961) obtained similar results, but varied the number of trials per problem within subjects rather than between subjects. These results are obtained by plotting performance against the number of trials, not the number of problems; the number of problems required to reach any specified performance level is reduced when the number of trials per problem is increased, within the specified limits.

*Kind of Stimuli.* Learning-set formation is faster when "stereometric" objects are used in the problems than when "pattern" stimuli are used (Chow, 1954; Harlow & Warren, 1952; Warren & Harlow, 1952), apparently because discrimination between stereometric objects is easier than discrimination between pattern stimuli (Harlow, 1945; Warren & Brookshire, 1959). (Pattern stimuli can differ in size, form, color, and pattern; stereometric objects usually differ on many additional dimensions. The distinction usually involves unidimensional versus multidimensional discriminanda.) Furthermore, there is more transfer from problems using pattern stimuli to problems using stereometric objects than from stereometric to pattern problems (Chow, 1954; see also Warren, 1953a, 1953b, 1953c).

*Contiguity of Stimulus and Locus of Response.* McDowell and Brown (1960) gave 234 "peripheral-cue" problems (with about a 3-inch separation between stimuli and response loci), and although there was significant within-problem improvement, there was no significant interproblem improvement. The subjects had had previous experience with standard learning-set problems, but apparently could not acquire a peripheral-cue learning set. Murphy and Miller (1955) gave one group of subjects standard learning-set problems and another group a kind of peripheral-cue problem in which the stimuli were located 6 inches above the manipulanda. The peripheral-cue group did not acquire a learning set, although they were given more than 600 problems; but the standard group acquired a learning set. After 576 problems the standard group was switched to the peripheral-cue problems, and the learning set disappeared; performance dropped to the level of the peripheral-cue group.

Riopelle, Wunderlich, and Francisco (1958) used pattern stimuli consisting of various numbers of concentric rings, with the relevant cue located in the center or in one of the rings. The performance level was higher, the closer the cue was to the perimeter; and although the number of rings (inversely related to size of stimuli) was a significant variable, there was little effect of the number of rings when the cue was on the perimeter. Warren (1963c) obtained similar results. He varied the positions of pattern stimuli on background cards and found the highest level of performance when the stimuli occupied the borders of the cards, the worst performance when the stimuli occupied the centers of the cards, and intermediate performance when the stimuli occupied the center and part of the border. The subjects in these studies probably displaced the stimulus cards by touching the edges of the cards, so that

the stimulus and locus of response were contiguous on the problems in which the cues were on the perimeters of the cards. Schuck, Polidora, McConnell, and Meyer (1961) found performance on center-cue problems superior to performance on border-cue problems; in their study the response units were centered on the stimuli (see also McConnell & Schuck, 1962).

In general, discrimination learning is easier when the stimuli and locus of response (reaching for food) are spatially contiguous than when they are separated (McClern & Harlow, 1954; Meyer, Polidora, & McConnell, 1961; Murphy & Miller, 1958; Schuck, 1960; Stollnitz & Schrier, 1962) and Otteson, Sheridan, and Meyer (1962) have concluded that:

monkeys, if given opportunities to do so, monitor their fingertips so closely that they fail to sample broadly from discriminanda. . . . Spatial contiguity becomes important only when the sampling process is affected [p. 937].

*Size of Stimuli.* Warren (1953b, 1953c) and Blazek and Harlow (1955) varied the size of pattern stimuli, by covering various percentages of the centers of the background cards, and found that increasing the area covered increased the performance level. According to Schuck et al. (1961), the data of such studies can be interpreted in terms of variation of the separation between stimulus and response loci. The greater the percentage of the card covered by the stimulus, the closer the stimulus is to the locus of response.

Two studies discussed above (Riopelle et al., 1958; Warren, 1953c) varied both the area and the position of the stimuli, and found significant effects of both variables. If increasing the area only reduces the separation between the stimulus and response locus, an interaction between area and position should be obtained; on the border-cue problems there is no separation, and therefore area should have no effect on performance; but on the center-cue problems, increas-

ing the area should reduce the separation between the stimulus and response locus, and therefore increase the level of performance. The required interaction was obtained in both studies, but was statistically significant only in the study by Riopelle et al. The required simple effects very closely approximated the data of Riopelle et al., but described Warren's data less well. Warren's curve for the border-cue position was not flat, but the rise was less than in the center-cue curve.

### *Performance Variables*

#### *Retention of Learned Associations.*

Several studies have shown remarkably high retention of responses to specific stimuli. In a study of learning-set development in infant monkeys, Mason, Blazek, and Harlow (1956) unintentionally repeated a series of 90 six-trial problems after an interval of about 1 month, and found more than chance correct Trial-1 responses in the repetition. Strong (1959) deliberately tested retention of discriminations in immature monkeys. The subjects learned each of 72 two-stimulus discriminations to criterion, and were given retention tests after each block of 12 problems and at intervals of from 30 to 210 days after the end of training. On the retention tests each pair of stimulus objects was presented once. The percentage correct responses on the retention tests ranged from 83% to 97%. Leary (1956) gave one-trial retention tests in which positive and negative stimuli were paired with new stimuli. Only 50% to 67% of the responses were correct, but the relatively low level of retention may be attributable to the kind of retention test used (see following section on Reactions to Novel Stimuli).

*Reactions to Novel Stimuli.* Experimentally naive monkeys tend to approach new stimuli and experimentally sophisticated monkeys tend to approach

familiar stimuli. Riopelle et al. (1962) gave experimentally naive monkeys four standard problems daily, and six daily problems each with one recurrent stimulus and one new stimulus. In the group of major relevance here, the six recurrent stimuli were positive on some problems and negative on other problems, the reward value changing randomly from day to day. This group exhibited a weak tendency to respond to the new stimuli on the first trial of the recurrent-stimulus problems for the first 20 days of the experiment, but eventually developed a tendency to respond to the familiar (recurrent) stimuli, although after 500 problems the preference for the familiar stimuli appeared to decline.

Fitzgerald and Davis (1960) gave a test trial on each problem, presenting either the positive or the negative stimulus with three new stimuli. When the positive stimulus was presented on test trials, all subjects tended to respond correctly. When the negative stimulus was presented on test trials, experimentally naive monkeys tended to respond to the position last occupied by the positive stimulus, and experimentally sophisticated monkeys responded most frequently to the negative stimulus. Experimentally sophisticated subjects, then, responded to the familiar, negative stimuli, but experimentally naive subjects did not.

Leary (1958) gave "ambiguous-cue" problems to subjects that had had experience with object discriminations and other kinds of problems for more than 3 years. The positive stimulus (A) and the "ambiguous" stimulus (C) were presented on the first trial (AC trial) of each ambiguous-cue problem, and the negative stimulus (B) and the ambiguous stimulus (C) were presented on the second trial (BC trial). When Stimulus A was chosen on the AC trial, subjects tended to choose Stimulus C on the BC trial, on which Stimulus B was a new

stimulus. There were many more errors on the AC trials (both stimuli new) than on the BC trials (B new, C correct). Furthermore, on transfer problems subjects initially tended to respond to Stimulus C, avoiding a new stimulus (D). These experimentally sophisticated subjects, then, tended to approach familiar stimuli.

Table 1 gives the designs and relevant data of five additional studies. As shown in Table 1, Riopelle (1955) gave five kinds of problems: (a) Positive-Positive—positive stimulus of a problem retained as positive stimulus in the next problem, (b) Negative-Negative—negative stimulus retained as negative, (c) Positive-Negative—positive stimulus retained as negative, (d) Negative-Positive—negative stimulus retained as positive (the reversals never occurred without at least one intervening problem), and (e) both stimuli new. The subjects were experimentally naive.

Performance on the Negative-Negative condition was superior to performance on the Positive-Positive condition, as indicated in Table 1, suggesting a tendency to approach the new stimulus. On the Negative-Negative condition the tendency to avoid the old negative stimulus and the assumed tendency of experimentally naive subjects to approach the novel stimulus are both correct; but on the Positive-Positive condition the tendency to approach the novel stimulus competes with the tendency to approach

the positive stimulus. The superiority of the Positive-Positive condition to the condition on which both stimuli were new indicates that the tendency to approach the old positive stimulus was stronger than the tendency to approach the new stimulus; on both conditions two approach tendencies compete, but there must have been less competition on the Positive-Positive condition. The superiority of the Negative-Positive condition to the Positive-Negative condition indicates that the tendency to approach the old positive stimulus was greater than the tendency to avoid the old negative stimulus (the subjects in the Negative-Positive condition overcame the tendency to avoid the old negative stimulus faster than the subjects in the Positive-Negative condition overcame the tendency to approach the old positive stimulus).

Brown, Overall, and Gentry (1958) gave Positive-Negative problems to one group, and Negative-Positive problems to another (reversals occurred without intervening problems). The subjects had had extensive laboratory experience including training on object-discrimination problems. First-trial performance on the Negative-Positive problems was superior to first-trial performance on the Positive-Negative problems (see Table 1).

According to Brown et al. (1958), "First-trial solution is explicable in terms of the formation of a concept of

TABLE 1  
ORDER OF EASE OF TRANSFER PROBLEMS FOLLOWING STANDARD DISCRIMINATION PROBLEMS

Experiment	Positive-Positive	Negative-Negative	Positive-Negative	Negative-Positive	Both New
Riopelle (1955)	2	1	5	4	3
Brown et al. (1958)	—	—	2	1	—
Brown et al. (1959)	—	1	2	—	—
Leary (1956)	1	2	—	—	—
Behar (1962)	1	2	—	—	—

Note.—Order is from easiest to most difficult. Labels are explained in text.

'newness' or 'novelty' and the association of a response with the conceptual stimulus class [p. 705]." The response tendency was to avoid new stimuli, and it might be noted that if subjects tend to avoid new stimuli, they may be said to have a "concept" of newness. (The subjects have a tendency to "avoid" a stimulus if the excitatory potential of the stimulus is less than the excitatory potential of the other stimulus.) Monkeys apparently can learn simple concepts (e.g., Hicks, 1956; Stone, 1961).

In the Negative-Positive condition one of the response tendencies, the assumed tendency of sophisticated subjects to avoid the new stimulus, is correct. As training continues and learning set develops, within-problem learning improves so that the subjects make very few errors and therefore acquire only a weak tendency to avoid the negative stimulus (as Brown et al., 1958, suggested). The weak tendency to avoid the negative stimulus is incorrect on the reversal, but does not compete effectively with the stronger, correct tendency to avoid the new stimulus. The reversal, then, is achieved without error.

Brown et al. (1958) included another group that was trained first on the Negative-Positive condition and then on the Positive-Negative condition. Performance on the second condition was superior to performance on the first. This finding is predictable from the stimulus-response analysis, since the excitatory potentials of the stimuli would be more nearly equal at the beginning of the second training period than the first.

Brown, Overall, and Blodgett (1959) gave Positive-Negative and Negative-Negative problems. All but 2 of the 33 subjects had been used by Brown et al. (1958) and the other 2 also had extensive laboratory experience on a variety of problems. First-trial performance on the Negative-Negative problems was

better than on the Positive-Negative problems.

In a Negative-Negative problem one of the response tendencies, the tendency to avoid the negative stimulus, is correct, and the problem should therefore be easier than the Positive-Negative reversal, in which both response tendencies are incorrect. The development of first-trial solution of both kinds of problem might indicate that a tendency to approach new stimuli gradually became stronger than the tendency to avoid new stimuli. (It might be noted that in learning to approach a given new stimulus, the subject does not necessarily learn an approach response to the class of "new stimuli," since experience makes a new stimulus familiar.) Alternatively the first-trial solutions could be explained by patterning. A pattern consisting of a familiar, positive stimulus and a new stimulus serves as a cue for a reversal; and a pattern consisting of a familiar, negative stimulus and a new stimulus does not. The assumption that the monkey can learn to respond differentially to such cues is supported by Riopelle and Copelan's (1954) finding that reversal can be stimulated by a change in background color.

Leary (1956) gave Positive-Positive and Negative-Negative problems to experimentally sophisticated subjects and found more errors on the latter, indicating that the subjects tended to avoid the new stimuli.

Using subjects that had previously acquired a learning set (Behar, 1961b), Behar (1962) gave problems with 4, 7, or 10 discrimination trials followed by a test trial on which either the positive or the negative stimulus was presented with a new stimulus. On the Positive-Positive test trials, about 96% of the responses were to the positive stimulus; and on the Negative-Negative test trials, 44% of the responses were to the negative stimulus, in line with Leary's (1956)

data. The tendency to avoid new stimuli need not be strong, since the tendency to avoid the *negative* stimulus is probably weak in subjects that have acquired a learning set. Only 2.6% of the responses of Behar's (1962) subjects on the discrimination trials (excluding Trial 1) were to the negative stimulus.

Additional data in Behar's (1962) study suggest that there is generalization from the positive to the negative stimulus within a problem. On the Negative-Negative test trials, the percentage of responses to the negative stimulus depended on whether or not errors occurred on the discrimination trials of the problem. When no errors were made on the discrimination trials, 63% of the responses were to the negative stimulus on test trials; but when one error or more occurred on the discrimination trials, the subjects responded to the negative stimulus on only 28% of the test trials. Furthermore, even when at least one error was made on the discrimination trials, the percentage of responses to the negative stimulus on test trials increased as the number of discrimination trials increased. After 4 discrimination trials, there were 23% responses to the negative stimulus; after 7, 28%; and after 10, 33%. As the tendency to approach the positive stimulus increased, with increasing training trials, the generalized tendency to approach the negative stimulus would increase, increasing the likelihood of responses to the negative stimulus on test trials.

Sheridan, Horel, and Meyer's (1962) data also support the conclusion. On each problem after the subject responded, either (a) both stimuli persisted for 2 seconds, (b) both stimuli disappeared immediately, (c) the positive stimulus alone persisted, or (d) the negative stimulus alone persisted. Performance was about the same on the first two conditions and inferior to performance on the last two conditions. Perform-

ance was about the same on the last two conditions (38% versus 35% errors). Generalization between the positive and negative stimuli is, then, apparently reduced when only one stimulus persists after the response, and therefore response competition is reduced and performance improves. (Mowrer's theory, 1960, Chapter 12, also accounts for these results. Generalization is not involved, but the probability of discriminating between the positive and negative stimuli is increased when only one stimulus persists, since it is "compared" with a stimulus trace, and learning is facilitated.)

The data of five other studies, included in Table 2, indicate that naive subjects tend to approach novel stimuli, but Table 2 also includes six studies which indicate that experimentally sophisticated subjects tend to approach novel stimuli. In all of these studies the discrimination trials were preceded by presentations of single stimuli, which minimizes within-problem generalization. Since the previously discussed studies used procedures more like the standard learning-set technique, with simultaneous presentation of the positive and negative stimuli (which enhances within-problem generalization), the development of a tendency to avoid novel stimuli must be related to within-problem generalization. Experimentally sophisticated subjects may retain a tendency to approach novel stimuli, but if generalization occurs from the positive to the negative stimulus within a problem, the tendency to approach familiar, negative stimuli becomes stronger than the tendency to approach novel stimuli, and the subjects avoid novel stimuli. If generalization is minimized, the subjects approach novel stimuli.

In the experimental conditions included in Table 2, each problem consisted of pretraining followed by discrimination trials with the pretraining

TABLE 2

ORDER OF EASE OF DISCRIMINATION PROBLEMS WITH SINGLE-STIMULUS PRETRAINING

Experiment	Reward-Positive	Nonreward-Negative	Reward-Negative	Nonreward-Positive
Warren and Brookshire (1959)	1		2	
Sinha (1958)	1		2	
Warren and Sinha (1959)	1		2	
Warren and Hall (1956)	1		2	
Blomquist and Harlow (1962)	2	1		
Brush et al. (1961)	2	1		
Mishkin et al. (1962)	2	1	3	4
Harlow and Hicks (1957)	2	1	1	2
Leary (1956)	2	1		
Moss and Harlow (1947)	2	1		
King and Harlow (1962)	2	1		
Group 75	1	2		
Group 50	2	1		
Group 25	2	1		

Note.—Order is from easiest to most difficult. Labels are explained in text.

stimulus and a new stimulus. In pre-training, the positive or negative stimulus of the discrimination trials was presented alone, with or without reward, yielding the four conditions given in Table 2.

After 200 pretraining trials, Warren and Brookshire's (1959) experimentally naive subjects made a mean of 6.0 errors to criterion on a Reward-Positive problem and 14.8 errors on a Reward-Negative problem. The tendency to approach the new stimulus appeared to be almost as strong as the conditioned-approach tendency resulting from 200 reinforcements, on the assumption that the subject attended to the stimulus on each pretraining trial. This assumption is, however, dubious, as indicated in the final section.

Sinha (1958) obtained data suggesting that there was an initial tendency to approach new stimuli in subjects that had served in one previous experiment. Eight pretraining trials were given in (a) the Reward-Positive condition and (b) the Reward-Negative condition. Sinha also gave (c) four baited presentations each of the positive stimulus and the negative

stimulus (Condition AB), and (d) no pretraining (control).

After 40 problems on each condition, the monkeys solved each kind of problem immediately (indicating that learning sets developed). Over the first 30 problems, however, performance on the Reward-Positive condition was superior to performance on the control condition, and performance on the control condition was superior to performance on the other two conditions. Performance on Condition AB was superior to performance on the Reward-Negative condition, but only over the first 10 problems. Sinha showed that the results are in line with Spence's (1938) stimulus-response analysis of discrimination learning until the learning sets are acquired. Since two competing approach tendencies were involved on each condition, the ordering must have resulted from differences in the amount of competition. The ordering of the conditions is explained if the tendency to approach new stimuli was weaker than the approach tendency resulting from four reinforcements.

Warren & Sinha (1959) obtained similar results, using experimentally

naive subjects and pattern stimuli (Sinha used stereometric objects). There was significant facilitation on Trial 1 of the Reward-Positive condition, and significant interference in the Reward-Negative condition; performance on Condition AB and on the control condition was not significantly different from chance. Warren and Hall (1956) also obtained essentially the same results, using subjects that were apparently relatively naive.

Blomquist and Harlow (1962) gave experimentally naive subjects problems each with one pretraining trial and one discrimination trial. There were more errors on Reward-Positive problems than on Nonreward-Negative problems, as expected if the subjects tend to approach new stimuli.

Experimentally sophisticated subjects in a study by Brush et al. (1961, Experiment II) avoided the pretraining object on the first discrimination trial of Nonreward-Negative and Nonreward-Positive problems 95% of the time, but on Reward-Positive and Reward-Negative problems they chose the pretraining object only 55% of the time. This finding would be expected if there was an initial tendency to approach new stimuli. Such a response tendency would compete with the tendency to approach the previously baited stimulus, but not with the tendency to avoid the previously nonbaited stimulus.

The fewest errors should be on the Nonreward-Negative condition, since both of the initial response tendencies are correct; the most errors should be on the Nonreward-Positive condition, since both initial response tendencies are incorrect; and there should be intermediate numbers of errors on the other two conditions, on each of which one response tendency is correct and one is incorrect. These postdictions were confirmed. Furthermore, there were fewer errors on the Reward-Positive condition

than on the Reward-Negative condition, indicating that the initial tendency to approach the new stimuli was relatively weak. When the tendency to approach the pretraining stimulus is stronger than the tendency to approach the new stimulus, there is less effective competition if the pretraining stimulus is correct than if it is incorrect.

Mishkin, Prockop, and Rosvold (1962) used the same subjects as Brush et al. (1961). Reward-Positive and Reward-Negative problems were given first, then Nonreward-Negative and Nonreward-Positive problems. Because of possible facilitating transfer from the first to the second series, this study does not allow a comparison of the first two conditions with the second two conditions. As in Brush et al., there were more errors on the Nonreward-Positive condition than on the Nonreward-Negative condition; but contrary to the earlier study, there were fewer errors on the Reward-Negative condition than on the Reward-Positive condition. The latter result would be expected if the tendency to approach the pretraining stimulus was weaker than the tendency to approach the new stimulus, exactly opposite to the conclusion suggested by the data of Brush et al. Mishkin et al., however, gave only one pretraining trial, and Brush et al. gave five. The tendency to approach the pretraining stimulus would be weaker in Mishkin et al., than in Brush et al., and could be weaker than the tendency to approach the new stimulus.

Performance on the Nonreward-Positive and Reward-Positive conditions in Mishkin et al. (1962) was better than chance, contrary to implications of the analysis outlined here. In all of the other studies except Sinha (1958), worse than chance performance was obtained when predicted by the analysis, and Sinha suggested that confounding stimu-

lus preferences were present in his study. Only Mishkin et al., in Table 2, used a between-subjects design permitting groups to acquire different learning sets. The tendency to approach new stimuli would enhance the effects of the learning set in the Nonreward-Negative group and the reversal learning set in the Reward-Negative group, but interfere with the effects of the reversal learning set of the Nonreward-Positive group and the learning set of the Reward-Positive group.

Harlow and Hicks (1957), Leary (1956), and Moss and Harlow (1947), all using experimentally sophisticated subjects, gave Reward-Positive and Nonreward-Negative problems, each with one or two pretraining trials. In all three studies performance on the Reward-Positive condition was inferior to performance on the Nonreward-Negative condition, as would be expected on the basis of the reasoning above if there was an initial tendency to approach new stimuli. (Moss & Harlow, 1947, also found that performance after one or two standard discrimination trials on which both stimuli were presented, with only the positive stimulus baited, was intermediate between the other two conditions, in agreement with this interpretation.)

Blomquist and Harlow (1962) obtained the same results in experimentally naive subjects, as indicated previously, but also found that the performance curves converged. The convergence indicates that the tendency to respond to new stimuli weakens as a result of prolonged training. (The weakening develops slowly; Blomquist and Harlow gave 360 problems on each condition; Harlow and Hicks gave only 90 on each condition; Moss and Harlow, 15; and Leary, 10.) On the discrimination trials of the Reward-Positive problems, the competition between the tendencies to approach

the positive stimulus and the new stimulus decreases as the latter tendency weakens, and performance improves. There is no competition on the Nonreward-Negative problems.

One further study shows that under certain conditions, within-problem generalization is not required for the development of a tendency to avoid novel stimuli. King and Harlow (1962) gave experimentally sophisticated subjects 720 problems, each with one pretraining and one discrimination trial. For one group 25% of the problems were Reward-Positive, for a second group 50% were Reward-Positive, and for a third group 75% were Reward-Positive. The other problems were Nonreward-Negative.

On the Reward-Positive problems, Group 75 was superior to Group 50, and Group 50 superior to Group 25. On the last blocks of Nonreward-Negative problems, Group 75 was inferior to Group 50, and Group 50 inferior to Group 25. In Group 75, performance on the Reward-Positive problems was superior to performance on the Nonreward-Negative problems; the difference was reversed in the other two groups and was greater in Group 25 than in Group 50.

The data suggest that a strong tendency to approach novel stimuli was developed in Group 25, in which the approach to novel stimuli was reinforced 75% of the time. Group 50 developed a weak tendency to approach novel stimuli (weaker than the tendency to approach the positive stimulus, since performance was better than chance on the Reward-Positive problems). Group 75 developed a weak tendency to avoid novel stimuli (weaker than the tendency to avoid the negative stimulus, since performance on the Nonreward-Negative problems was better than chance). The approach response to novel stimuli

was nonreinforced 75% of the time in Group 75.

In summary, subjects have a fairly weak tendency to approach novel stimuli (except when the percentage reinforcement of this tendency is large, as in King and Harlow's 25% group); but if within-problem generalization occurs from the positive to the negative stimulus, a generalized tendency to approach the negative stimulus becomes stronger than the tendency to approach novel stimuli, and subjects begin to approach familiar stimuli. Under conditions which minimize within-problem generalization, the tendency to approach novel stimuli is gradually extinguished by prolonged training in which the tendency is 50% reinforced.

*Transfer of Learning Sets.* Learning sets transfer and may interfere with the acquisition of a different learning set. Shifting the subjects from one kind of problem to another brings about a large number of errors, but as the shifts continue, progressively fewer errors occur (Harlow, 1959). Warren and Sinha (1959) gave monkeys three series of discrimination learning-set problems. Between the first and second series, the monkeys served in studies of discrimination, discrimination reversal, and conditional discrimination. Between the second and third series, they were given double-alternation training. Performance improved from the first to the second series, but deteriorated from the second to the third.

Behar (1961a) found that experimentally sophisticated monkeys transferred an object-discrimination learning set which interfered with the formation of an object-alternation learning set, and Levine and Harlow (1959) found that training on object-discrimination problems interfered with later oddity learning.

Harlow (1949) gave position-discrimi-

nation and object-discrimination problems to monkeys that had previously acquired an object-discrimination learning set and a discrimination-reversal learning set. By the end of training the monkeys could solve problems of both kinds fairly rapidly, but not as rapidly as when they had been given problems of only one kind. The position discriminations were solved less rapidly than the object discriminations. There was, then, interference between the learning sets, and the object-discrimination learning set interfered more on the position problems than the position learning set interfered with the object discriminations. Warren (1959) obtained essentially the same results using monkeys that had acquired an object-discrimination learning set in a previous study (Sinha, 1958).

Riopelle (1953) gave five problems with new stimuli each day, and one daily reversal problem, for 63 days. In the last block of 14 days, the percentage correct responses on Trial 2 of the new problems were greater than the percentage correct responses on Trial 2 of the reversal problems, indicating that the discrimination learning set continued to interfere with the reversal learning set.

One other study (Riopelle & Chinn, 1961) relevant to this conclusion showed that monkeys can learn a position habit and a learning set simultaneously, but there is interference between these response tendencies. On the first trial of every problem the correct stimulus was on the same side. Over the last block of 50 problems the subjects gave 85% correct responses on the first trial, but a slightly smaller percentage correct responses on the second trial; on half of the second trials the correct stimulus had changed position. In the last 200 problems there were 12% incorrect responses on Trials 5 and 6, and 85% of these errors were made on the side of the position habit.

*Systematic Patterns of Responses.* Systematic patterns of responses occur, not only after a problem has been mastered, when all responses are correct, but also in the "presolution" period. The systematic response tendencies occurring in the presolution period have been called hypotheses (Krechevsky, 1932; Levine, 1959), strategies (Bowman, 1961), and error factors (Harlow, 1949, 1950, 1959). The studies summarized below used Harlow's error factor model (see Levine, 1959, for applications of the "hypothesis" model).

1. *Position Preference:* None of Harlow's (1950) 11 subjects had a position preference in the first hundred problems, although 3 had significant position preferences in the second hundred problems, and Harlow (1959) has concluded that position preference is an "essentially unimportant" error factor in rhesus monkeys.

2. *Position Alternation:* Spontaneous position alternation apparently does not occur in monkeys, or at most seldom occurs (see Behar, 1961b; Warren & Sinha, 1959).

3. *Stimulus Perseveration (Stimulus Preference):* Stimulus perseveration decreases as training progresses (Harlow, 1950). As training progresses, the number of stimulus-perseveration error sequences decreases, and the duration of the error sequences decreases (Harlow, 1959). Stimulus perseveration also appears to be stronger in young monkeys than in mature animals (Harlow, 1959).

Brush et al. (1961) and Harlow (1959) have listed stimulus characteristics that apparently are associated with spontaneous approach and avoidance tendencies, and have suggested that generalization from previously encountered positive and negative stimuli also affects stimulus perseveration.

4. *Stimulus Alternation:* Stimulus alternation apparently has a low proba-

bility of occurrence in monkeys, but may be related to the Response-Shift Errors, discussed below. Behar (1961a) studied stimulus alternation, but in monkeys that had previously acquired a discrimination learning set.

5. *Differential-Cue Error:* The differential-cue error is the frequency of errors on the first trial on which the correct stimulus object changes position (differential-cue trial), relative to errors on trials on which the correct stimulus remains in the same position (multiple-cue trials). The differential-cue error persists longer than stimulus perseveration and is a function of the number of multiple-cue trials preceding the first differential-cue trial (Harlow, 1959). Harlow (1950) found almost as great a differential-cue error in the second hundred six-trial problems as in the first hundred.

The first-trial outcome does not seem to affect the differential-cue error; Harlow and Warren (1952) found about the same strength of differential-cue error following a first-trial error as following a first-trial correct response.

Finally, it has been shown that electroshock convulsions increase the strength of the differential-cue error. After a learning set had been acquired, the subjects in a study by Braun, Patton, and Barnes (1952) were given 20 electroshock convulsions, followed by further learning-set problems. On the postshock problems, the subjects tended to persevere the response to the position last rewarded (differential-cue error).

6. *Response-Shift Errors:* At least two kinds of response-shift error have been described. According to one definition, response shift is a greater frequency of errors following an initial correct response than following an initial error. A second definition describes response shift as the number of errors following

a series of correct responses when the initial response was correct, relative to the number of such errors when the initial response was incorrect. It has been assumed that the response-shift error results from "a strong tendency of the monkey to respond to both stimuli in the object-discrimination learning situation [Harlow, 1959, p. 516]."

Harlow (1959) has concluded that response shift is the most persistent error factor. Harlow (1950) found that the percentage response-shift errors increased from 29% in the first hundred 6-trial problems to 57% in the second hundred.

Although Harlow suggested in 1950 that Hull's (1943) construct "reactive inhibition" might account for response-shift errors, he had abandoned that possible explanation by 1959, since "Davis (1956) found little evidence that work inhibition is either an important or consistent variable influencing performance of rhesus monkeys [p. 518]." The relevance of Davis' results is debatable, since his results could be attributed to competition between the inhibitory effect of work (reactive inhibition) and a facilitating effect of work on "attention" or some other orienting-response mechanism (increasing work sometimes increased the performance level). Harlow (1959) has suggested that response shift may be related to spontaneous alternation, such as is found in rats, but spontaneous position alternation and stimulus alternation seem to be infrequent in monkeys.

Early in training there are more correct responses following a correct response on Trial 1 than following an error on Trial 1; but later the trend is reversed—there are more correct responses following an error than following a correct first-trial response. Riopelle (1953) found the former trend in the first 105 of 315 problems, and the latter in the last

105 problems. Davis (1961) obtained only the former trend in his subjects, but they were given only 140 problems.

Harlow and Warren (1952) gave 450 problems with pattern stimuli, then 144 problems with stereometric stimuli, and found the former trend in both series of problems. McConnell and Schuck's (1962) subjects also exhibited the early trend throughout training on 832 problems with pattern stimuli, but there was only minimal within-problem learning following an error, indicating the presence of strong stimulus preferences.

Harlow and Hicks (1957), using relatively naive subjects, designated the object chosen on the first trial as correct or incorrect on different problems by baiting both stimuli or neither stimulus on the first trial. (They also included two other conditions, discussed previously.) There were more errors on the nonbaited problems than on the baited problems in the first block of problems (15 problems per condition), and about equal numbers of errors on the two conditions over the five remaining blocks, in line with the conclusion. Brush et al. (1961) also used relatively naive subjects and the same kind of control of first-trial reward. Over the first block of problems (20 problems per condition) there were more errors on the nonbaited than on the baited problems, and thereafter there was little difference.

Behar (1961b) found both trends in relatively naive monkeys. The curve depicting percentage correct responses after an initial error crossed the curve for percentage correct after an initial correct response between the fourth and fifth blocks of 50 problems. Another study by Behar (1962) suggests that the effect may eventually disappear. Using experimentally sophisticated subjects (previously trained by Behar, 1961b), he found 91.6% correct responses on Trial 2 following a correct Trial-1 re-

sponse, and 93.5% correct following an initial error. Warren (1959) also used experimentally sophisticated subjects, and found that performance was not significantly different following first-trial errors and first-trial correct responses. The subjects were given only 16 problems, however, and were trained to criterion on each one.

### THEORETICAL ANALYSIS

The term "learning set" has usually been used to refer to one-trial learning or a close approach to one-trial learning occurring after training on problems that have a common basis of solution (such as "approach odd stimulus" in oddity problems). The theoretical analysis of one-trial learning outlined below is based on Hull-Spence theory.

#### *Basic Assumptions*

One-trial learning can occur if (a) an increment in habit strength produced by a single reinforcement or an increment in inhibitory strength produced by a single nonreinforcement makes the excitatory potential of the positive stimulus greater than that of the negative stimulus, and (b) the difference between the excitatory potentials of the positive and negative stimuli is greater than oscillatory inhibition. Spence (1956, pp. 100-101) has assumed that the range of oscillatory inhibition is reduced as a result of prolonged practice in the instrumental conditioning situation, and it seems reasonable to assume that the reduction would also occur in the learning-set situation. To satisfy Condition *b* the reduced oscillatory inhibition would have to be smaller than single increments in habit and inhibition.

Condition *a* requires that the positive and negative stimuli have about equal habit strengths and about equal inhibitory strengths before the first response on a problem. This would occur, theo-

retically, if the subject experienced a large number of stimulus elements on many dimensions (as when multidimensional and "junk" stimuli are used), and responses to these elements were about equally often rewarded and nonrewarded over a long series of problems. As a result of generalization, unused stimulus elements on these dimensions would also acquire habit and inhibitory strengths of magnitudes about equal to those of stimulus elements that were directly experienced. Generalization between multidimensional stimuli could be attributed to a failure of discrimination between the stimuli, resulting from similarity of common stimulus elements, or to "primary stimulus generalization" among the similar stimulus elements. It would be less reasonable to attribute the generalization to identical stimulus elements.

It might be argued that the number of problems actually required to attain the learning set is too small for the equalization of habit strengths and of inhibitory strengths, since not enough stimulus elements could be experienced, but Riopelle and Rogers (1956) found more generalization in college students given several conditioned stimuli than in subjects given a constant conditioned stimulus. White and Spiker (1960) obtained the same result in preschool children, and White (1961) obtained data in line with this finding in a discrimination learning task, again using preschool children. In a relatively short series of training problems, the subject may receive practice with several members of each stimulus dimension, and therefore generalize widely along each dimension. (However, Zimmerman, 1962, found that neither multiple-problem training nor overlearning increased generalization in infant monkeys.)

#### *Analyses of Specific Findings*

*Learning-Set Acquisition Curves.* Initially, the curves should be positively ac-

celerated, since associations are learned to a relatively large number of new stimulus elements in each of the early problems. Later in the course of learning-set formation, as relatively fewer new stimulus elements are added on each new problem, the acquisition curve should be negatively accelerated. In general, the acquisition curve seems to be ogival (Harlow, 1959).

The theory requires that the subject be given practice on a large number of problems, and as indicated above, the data seem to satisfy this requirement. Generally, well over 100 problems must be given before there is evidence that a learning set has been acquired.

*Experimental Variables.* The relation between within-problem learning and problems to criterion is obtained only if the maximum number of trials on each problem is some number greater than 12 and less than 50. Evidence from studies of the solution of a single two-stimulus problem by monkeys suggests that this maximum number of trials may be the point in training at which overlearning begins. The theory outlined here requires that overlearning interfere with the formation of a learning set, because overlearning leads to stimulus preferences which would have to be overcome by the development of equally strong preferences for other stimuli before the learning set would be evidenced. The habit strengths of all stimulus elements (and their inhibitory strengths) must be approximately equal for one-trial learning to occur. Overlearning would raise the magnitude at which the habits would be equal, and therefore increase the number of training problems required to equalize the habits.

The analysis also requires, however, that there be some lower limit to the magnitude at which the habits are equalized. Otherwise, fewer problems would be required for the development

of one-trial learning, the fewer the trials per problem (since weaker stimulus preferences would develop). The assumption of a lower limit is reasonable, since it is known that "spontaneous" preferences occur; at least the majority of these stimulus preferences must be overcome before one-trial learning can predominate in the subject's performance. (It may be noted that this reasoning requires the further assumption that most spontaneous preferences be "learned" but not "overlearned.") Therefore, the theory predicts a decrease in the number of problems to criterion as the number of trials per problem increases to some number between 12 and 50, and an increase in problems to criterion as the number of trials per problem increases beyond this limit.

The slower learning-set formation with pattern stimuli than with stereometric stimuli reflects the greater difficulty of the pattern problems than the stereometric problems. The difficulty of the pattern problems may result from generalization between the positive and negative stimuli of a problem. This explanation would require the assumption that training on a large number of problems either reduces the generalization or reduces the oscillatory inhibition to a value smaller than the difference between the conditioned and the generalized habit or inhibition (the latter alternative seems more reasonable, since it is only an extension of the assumption regarding the reduction of oscillatory inhibition with prolonged training, and since the former alternative appears to be incompatible with the conclusion about within-problem generalization).

An alternative explanation is also in line with Hull-Spence theory. Only effective stimuli can acquire habit strength; to be "effective," stimuli must be observed at the moment of responding (Hull, 1943; Spence, 1956). Even

naive monkeys probably have appropriate observing responses for stereometric objects, since their history includes everyday experience with such objects, but monkeys usually have less pre-experimental experience with patterns. During prolonged training, however, appropriate observing responses are acquired and eventually a learning set develops.

The orienting response mechanism also explains the effect of the spatial contiguity of the stimuli and response loci. The probability of observing the relevant stimuli in the noncontiguous condition is low, and therefore the probability of observing the relevant stimulus differences is even lower (Stollnitz & Schrier, 1962).

The increase in the rate of learning-set formation with increase in the size of pattern stimuli (percentage of background cards covered by stimuli) can be explained by the same mechanism, as noted in the review section.

#### *Retention of Learned Associations.*

The theory requires that learned associations be retained, but leads to the prediction that retention should be greater early in learning than later, when generalization is extensive. To avoid the complication of responses to novel stimuli in the tests of retention, reversal problems could be used. The studies of reversal learning in successive problems have uniformly confirmed the prediction, since performance on the reversal problems gradually improves until late in training one-trial reversals occur (e.g., Harlow, 1949; Riopelle, 1953). Furthermore, experimentally naive monkeys probably acquire a reversal learning set more slowly than monkeys that have previously acquired a discrimination learning set (Harlow, 1949). The analysis outlined here also predicts that reversal should be faster when the first reversal problem experienced by the sub-

ject is given late in discrimination learning-set training than when it is given early.

*Within-Problem Generalization.* There is generalization from the positive to the negative stimulus within a problem in subjects that have had prolonged training. The direction of the generalization results from the greater frequency of correct responses in a problem in such subjects; very few errors occur after the first trial of a problem. The increase in habit in a problem is therefore greater than the increase in inhibition, and the habit generalized from the positive to the negative stimulus is apparently greater than the inhibition developed.

Some of the elements of the negative stimulus may be similar to elements of the positive stimulus, providing a basis for generalization, but temporal contiguity is apparently also required.

*Reactions to Novel Stimuli.* The tendency of experimentally naive monkeys to approach novel stimuli may be associated with a "curiosity drive" (Harlow, 1953), or more accurately with the drive stimuli produced by such a motive, but the source of the tendency is not relevant here. After prolonged training, no stimuli are entirely new, since at least some of their elements have been previously experienced, and therefore the tendency to approach new stimuli should eventually disappear. The response tendencies of sophisticated monkeys are discussed in the review section.

*Transfer of Learning Sets.* A learning set should transfer between problems if the contexts in which the stimuli are presented are identical and if the stimuli are of the same kind. Changing the stimulus context can change the response tendencies, and the response tendencies would be different if the stimuli used in one kind of problem had few elements in common with those of the other kind of problem. If transfer occurs, it may

result in interference or in facilitation, depending on the kinds of problem. In most cases, however, facilitation and interference probably result from the transfer of performance sets, especially the transfer of orienting responses, rather than of habit and inhibition. It seems likely that one kind of orienting response is acquired on position-alternation problems, for example, and another kind on stimulus-discrimination problems.

*Systematic Patterns of Responses.* With the exception of the response-shift error, all of the error factors can be explained on the basis of spontaneous or previously acquired response tendencies to specific stimulus elements. The differential-cue error, for example, may be related to position habits, and should persist longer than such an error factor as stimulus perseveration, since the response to position is rewarded on each multiple-cue trial.

Regarding the response-shift error, it has been shown that early in training there are more correct responses following a correct response on Trial 1 than following a first-trial error, but that this trend is reversed later in training. Behar (1961b) has concluded that:

The negative reinforcement effect of nonreward is absent in naive animals but is acquired in successful discrimination learning [p. 51].

Spence (1960) has assumed that in instrumental reward conditioning,

the inhibitory effect of a non-reinforcement is related to the strength of the response tendency and would be zero at very low levels, such as presumably would be the case in the earliest stages of learning [p. 98].

Riopelle and Francisco's (1955) data support this assumption (see also Riopelle, Francisco, & Ades, 1954, who used somewhat different conditions, but obtained similar results). One of their groups was trained on problems with neither stimulus baited on the first trial

(nonbaited group) and another with both stimuli baited on the first trial (baited group). The stimulus that the subject responded to on the first trial was baited on the remaining trials and the other stimulus was nonbaited. After 250 training problems, the subjects were given 50 transfer problems, each with a predetermined positive stimulus baited on all trials. On the transfer problems, the nonbaited group repeated responses 86% of the time following an initial non-reward and the baited group only 45%.

Riopelle et al. (1962) gave three groups of experimentally naive subjects four daily problems with new stimuli, and six daily problems each with one new stimulus and one recurrent stimulus. The recurrent stimuli were always positive in one group, always negative in another, and sometimes positive and other times negative in the third group. On the "new-stimulus" problems, the "recurrent-positive" group gave more correct responses following a correct first-trial response than following an initial error throughout training (except in the fourth block of 10 days). In the "recurrent-negative" group this trend was obtained over the first third of training (20 days), but the opposite trend was obtained over the last 40 days. These results confirm the general conclusion of this section, on the assumption that the negative reinforcement effect of a non-reward was acquired sooner and was stronger in the recurrent-negative group than in the other two groups.

The increment in inhibition resulting from a single nonreward increases as training continues; but the increment in habit strength, resulting from a single reinforcement, is a negatively accelerated function of habit strength before the reinforcement and would be larger early in training than later. Early in training, therefore, an increment in habit is greater than an increment in inhibi-

tion; and a single reward has a greater effect on performance than a single nonreward. Later in training, the increment in habit is smaller than the increment in inhibition, and a nonreward has more influence on performance than a reward.

One might speculate that if habit and inhibition ever became asymptotic in value, performance could not rise above the chance level, since all stimulus elements would have the maximum possible strengths of habit and inhibition. (Because of oscillatory inhibition, responses would occur, but no further learning could occur.) Furthermore, the performance level should decline as the approach to the asymptote becomes close. McConnell and Schuck (1962) gave 832 six-trial problems with pattern stimuli, and found such a decline from the twelfth to the thirteenth block of 64 problems, except in a slow-learning subgroup (see also Schuck et al., 1961). However, most of the existing data on monkeys with extensive learning-set experience suggest that the prediction is incorrect, and if further research fails to confirm the prediction, the assumption that habit and inhibition are permanent will have to be modified. (It may be possible to use such data to quantify the decay of habit and inhibition over time.) It should perhaps be emphasized that the tests of this prediction are not crucial, since the other derivations from the theory require only "relative permanence" of habit and inhibition.

#### REFERENCES

- BEHAR, I. Analysis of object-alternation learning in rhesus monkeys. *J. comp. physiol. Psychol.*, 1961, 54, 539-542. (a)
- BEHAR, I. Learned avoidance of nonreward. *Psychol. Rep.*, 1961, 9, 43-52. (b)
- BEHAR, I. Evaluation of the significance of positive and negative cue in discrimination learning. *J. comp. physiol. Psychol.*, 1962, 55, 502-504.
- BLAZEK, N. C., & HARLOW, H. F. Persistence of performance differences on discriminations of varying difficulty. *J. comp. physiol. Psychol.*, 1955, 48, 86-89.
- BLOMQUIST, A. J., & HARLOW, H. F. The effects of reward and nonreward on discrimination learning by monkeys. Unpublished manuscript, University of Wisconsin, 1962.
- BOWMAN, R. E. Discrimination learning set under intermittent and secondary reinforcement. Paper read at American Psychological Association, New York, September 1961.
- BRAUN, H. W., PATTON, R. A., & BARNES, H. W. Effects of electroshock convulsions upon the learning performance of monkeys: I. Object-quality discrimination learning. *J. comp. physiol. Psychol.*, 1952, 45, 231-238.
- BROWN, W. L., OVERALL, J. E., & BLODGETT, H. C. Novelty learning sets in rhesus monkeys. *J. comp. physiol. Psychol.*, 1959, 52, 330-332.
- BROWN, W. L., OVERALL, J. E., & GENTRY, G. V. Conceptual discrimination in rhesus monkeys. *J. comp. physiol. Psychol.*, 1958, 51, 701-705.
- BRUSH, E. S., MISHKIN, M., & ROSVOLD, H. E. Effects of object preferences and aversions on discrimination learning in monkeys with frontal lesions. *J. comp. physiol. Psychol.*, 1961, 54, 319-325.
- BUGELSKI, B. R. *The psychology of learning*. New York: Holt, 1956.
- CHOW, K. L. Effects of temporal neocortical ablation on visual discrimination learning sets in monkeys. *J. comp. physiol. Psychol.*, 1954, 47, 194-198.
- DAVIS, R. T. Problem-solving behavior of monkeys as a function of work variables. *J. comp. physiol. Psychol.*, 1956, 49, 499-506.
- DAVIS, R. T. Discriminative performance of monkeys irradiated with X-rays. *Amer. J. Psychol.*, 1961, 74, 86-89.
- FITZGERALD, R. D., & DAVIS, R. T. The role of preference and reward in the selection of discriminanda by naive and sophisticated rhesus monkeys. *J. genet. Psychol.*, 1960, 97, 227-235.
- HARLOW, H. F. Studies in discrimination learning by monkeys: III. Factors influencing the facility of solution of discrimination problems by rhesus monkeys. *J. gen. Psychol.*, 1945, 32, 213-227.
- HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, 56, 51-65.
- HARLOW, H. F. Analysis of discrimination learning by monkeys. *J. exp. Psychol.*, 1950, 40, 26-39.
- HARLOW, H. F. Motivation as a factor in the acquisition of new responses. In, *Current theory and research in motivation: A sym-*

- posium. Lincoln: Univer. Nebraska Press, 1953.
- HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.
- HARLOW, H. F., & HICKS, L. H. Discrimination learning theory: Uniprocess vs. duoprocess. *Psychol. Rev.*, 1957, 64, 104-109.
- HARLOW, H. F., & WARREN, J. M. Formation and transfer of discrimination learning sets. *J. comp. physiol. Psychol.*, 1952, 45, 482-489.
- HICKS, L. H. An analysis of number-concept formation in the rhesus monkey. *J. comp. physiol. Psychol.*, 1956, 49, 212-218.
- HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
- KING, J. E., & HARLOW, H. F. Effects of ratio of Trial 1 reward to nonreward on the discrimination learning of macaque monkeys. *J. comp. physiol. Psychol.*, 1962, 55, 872-875.
- KRECHEVSKY, I. "Hypothesis" versus "chance" in the presolution period in sensory discrimination learning. *U. Calif. Publ. Psychol.*, 1932, 6, 27-44.
- LEARY, R. W. The rewarded, and unrewarded, the chosen, and the unchosen. *Psychol. Rep.*, 1956, 2, 91-97.
- LEARY, R. W. The learning of ambiguous cue-problems by monkeys. *Amer. J. Psychol.*, 1958, 71, 718-724.
- LEVINE, M. A model of hypothesis behavior in discrimination learning set. *Psychol. Rev.*, 1959, 66, 353-366.
- LEVINE, M., & HARLOW, H. F. Learning-sets with one- and twelve-trial oddity-problems. *Amer. J. Psychol.*, 1959, 72, 253-257.
- LEVINE, M., HARLOW, H. F., & PONTRELLI, T. Supplementary report: The effects of problem length on transfer during learning-set performance. *J. exp. Psychol.*, 1961, 61, 192.
- LEVINE, M., LEVINSON, B., & HARLOW, H. F. Trials per problem as a variable in the acquisition of discrimination learning set. *J. comp. physiol. Psychol.*, 1959, 52, 396-398.
- MCCLEARN, G. E., & HARLOW, H. F. The effect of spatial contiguity on discrimination learning by rhesus monkeys. *J. comp. physiol. Psychol.*, 1954, 47, 391-394.
- MCCONNELL, D. G., & SCHUCK, J. R. An analysis of primate responses to colored filmstrip stimuli. *J. comp. physiol. Psychol.*, 1962, 55, 505-510.
- MCDOWELL, A. A., & BROWN, W. L. Peripheral cue learning set in rhesus monkeys. *J. genet. Psychol.*, 1960, 96, 129-132.
- MASON, W. A., BLAZEK, N. C., & HARLOW, H. F. Learning capacities of the infant rhesus monkey. *J. comp. physiol. Psychol.*, 1956, 49, 449-453.
- MEYER, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, 41, 10-16.
- MEYER, D. R., & HARLOW, H. F. The development of transfer of response to patterning by monkeys. *J. comp. physiol. Psychol.*, 1949, 42, 454-462.
- MEYER, D. R., POLIDORA, V. J., & MCCONNELL, D. G. Effects of spatial S-R contiguity and response delay upon discriminative performances by monkeys. *J. comp. physiol. Psychol.*, 1961, 54, 175-177.
- MISHKIN, M., PROCKOP, E. S., & ROSVOLD, H. E. One-trial object-discrimination learning in monkeys with frontal lesions. *J. comp. physiol. Psychol.*, 1962, 55, 178-181.
- MOON, L. E., & HARLOW, H. F. Analysis of oddity learning by rhesus monkeys. *J. comp. physiol. Psychol.*, 1955, 48, 188-195.
- MOSS, E. M., & HARLOW, H. F. The role of reward in discrimination learning in monkeys. *J. comp. physiol. Psychol.*, 1947, 40, 333-342.
- MOWRER, O. H. *Learning theory and behavior*. New York: Wiley, 1960.
- MURPHY, J. V., & MILLER, R. E. The effect of spatial contiguity of cue and reward in the object-quality learning of rhesus monkeys. *J. comp. physiol. Psychol.*, 1955, 48, 221-224.
- MURPHY, J. V., & MILLER, R. E. Effect of the spatial relationship between cue, reward, and response in simple discrimination learning. *J. exp. Psychol.*, 1958, 56, 26-31.
- OTTESON, M. I., SHERIDAN, C. L., & MEYER, D. R. Effects of stimulus-response isolation on primate pattern discrimination learning. *J. comp. physiol. Psychol.*, 1962, 55, 935-938.
- RESTLE, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, 65, 77-91.
- RESTLE, F. Note on the "hypothesis" theory of discrimination learning. *Psychol. Rep.*, 1960, 7, 194.
- RESTLE, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, 69, 329-343.
- RIOPELLE, A. J. Transfer suppression and learning sets. *J. comp. physiol. Psychol.*, 1953, 46, 108-114.
- RIOPELLE, A. J. Learning sets from minimum stimuli. *J. exp. Psychol.*, 1955, 49, 28-32.
- RIOPELLE, A. J., & CHINN, R. McC. Position habits and discrimination learning by monkeys. *J. comp. physiol. Psychol.*, 1961, 54, 178-180.
- RIOPELLE, A. J., & COPELAN, E. L. Discrimination reversal to a sign. *J. exp. Psychol.*, 1954, 48, 143-145.

- RIOPELLE, A. J., CRONHOLM, J. N., & ADDISON, R. G. Stimulus familiarity and multiple discrimination learning. *J. comp. physiol. Psychol.*, 1962, 55, 274-278.
- RIOPELLE, A. J., & FRANCISCO, E. W. Discrimination learning performance under different first-trial procedures. *J. comp. physiol. Psychol.*, 1955, 48, 90-93.
- RIOPELLE, A. J., FRANCISCO, E. W., & ADES, H. W. Differential first-trial procedures and discrimination learning performance. *J. comp. physiol. Psychol.*, 1954, 47, 293-297.
- RIOPELLE, A. J., & ROGERS, J. P. Transfer of training to orthogonal dimensions. *J. exp. Psychol.*, 1956, 52, 367-370.
- RIOPELLE, A. J., WUNDERLICH, R. A., & FRANCISCO, E. W. Discrimination of concentric-ring patterns by monkeys. *J. comp. physiol. Psychol.*, 1958, 51, 622-626.
- SCHUCK, J. R. Pattern discrimination and visual sampling by the monkey. *J. comp. physiol. Psychol.*, 1960, 53, 251-255.
- SCHUCK, J. R., POLIDORA, V. J., MCCONNELL, D. G., & MEYER, D. R. Response location as a factor in primate discrimination. *J. comp. physiol. Psychol.*, 1961, 54, 543-545.
- SHERIDAN, C. L., HOREL, J. A., & MEYER, D. R. Effects of response-induced stimulus change on primate discrimination learning. *J. comp. physiol. Psychol.*, 1962, 55, 511-514.
- SINHA, M. M. Effects of reinforcing single stimuli upon subsequent discrimination learning by rhesus monkeys. *J. comp. physiol. Psychol.*, 1958, 51, 87-91.
- SPENCE, K. W. Analysis of the formation of discrimination habits in chimpanzees. *J. comp. Psychol.*, 1938, 23, 77-100.
- SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.
- SPENCE, K. W. *Behavior theory and learning*. Englewood Cliffs, N. J.: Prentice-Hall, 1960.
- STOLLNITZ, F., & SCHRIER, A. M. Discrimination learning by monkeys with spatial separation of cue and response. *J. comp. physiol. Psychol.*, 1962, 55, 876-881.
- STONE, G. C. Attainment of color, form, and size concepts by rhesus monkeys. *J. comp. physiol. Psychol.*, 1961, 54, 38-42.
- STRONG, P. N., JR. Memory for object discriminations in the rhesus monkey. *J. comp. physiol. Psychol.*, 1959, 52, 333-335.
- WARREN, J. M. Additivity of cues in visual pattern discriminations by monkeys. *J. comp. physiol. Psychol.*, 1953, 46, 484-486. (a)
- WARREN, J. M. Effect of geometrical regularity on visual form discrimination by monkeys. *J. comp. physiol. Psychol.*, 1953, 46, 237-240. (b)
- WARREN, J. M. The influence of area and arrangement on visual pattern discrimination by monkeys. *J. comp. physiol. Psychol.*, 1953, 46, 231-236. (c)
- WARREN, J. M. Solution of object and positional discriminations by rhesus monkeys. *J. comp. physiol. Psychol.*, 1959, 52, 92-93.
- WARREN, J. M., & BROOKSHIRE, K. H. Stimulus generalization and discrimination learning by primates. *J. exp. Psychol.*, 1959, 58, 348-351.
- WARREN, J. M., & HALL, J. F. Discrimination of visual patterns as a function of motivation and frequency of reinforcement. *J. genet. Psychol.*, 1956, 88, 245-249.
- WARREN, J. M., & HARLOW, H. F. Learned discrimination performance by monkeys after prolonged postoperative recovery from large cortical lesions. *J. comp. physiol. Psychol.*, 1952, 45, 119-126.
- WARREN, J. M., & SINHA, M. M. Interactions between learning sets in monkeys. *J. genet. Psychol.*, 1959, 95, 19-25.
- WHITE, S. H. Effects of pretraining with varied stimuli on children's discrimination learning. *Child Develpm.*, 1961, 32, 745-753.
- WHITE, S. H., & SPIKER, C. C. The effect of a variable conditioned stimulus upon the generalization of an instrumental response. *Child Develpm.*, 1960, 31, 313-319.
- ZIMMERMAN, R. R. Form generalization in the infant monkey. *J. comp. physiol. Psychol.*, 1962, 55, 918-923.

(Received March 6, 1963)

# DELAY OF REINFORCEMENT:

## A HISTORICAL REVIEW<sup>1</sup>

K. EDWARD RENNER

*University of Pennsylvania*

The independent variable of delay of reinforcement is traced from its early empirical history through the systematic treatment provided by Hull and Spence. The empirical findings and theoretical implications of recent experimental studies are noted. The effects of delay of reward and delay of punishment on the learning process are considered, the comparative studies with human Ss are reported, and the relevance of delay of reinforcement to abnormal behavior and personality is summarized.

The relevance of delay of reinforcement to an increasing number of theoretical issues and the need to provide a comprehensive review of the historical context into which these recent studies can be viewed is one purpose of this paper. A second purpose is to review the current literature critically with the aim of clarifying the problems needing additional experimental work, and finally, to indicate the relevance of delay of reinforcement to several aspects of more complex human behavior.

### HISTORICAL BACKGROUND

#### *On the Nature of Learning*

A comparative study by Hunter (1913) apparently was the first experimental study on delay of reinforcement, although the theoretical significance of delay of reinforcement as an independent variable can be traced to Watson's (1917) experiment on delayed reward of a digging response. Watson found that a 30-second delay had no effect on the acquisition of the response and concluded that his results presented a difficult problem for the view that the get-

ting of food stamps in the most recent act.

Warden and Haas (1927), after an earlier study by Yarbrough (1921), followed with an extension of Watson's (1917) study; subjects were run in an 8-unit maze rather than a problem box and feeding in the goal box was permitted after 1- or 5-minute delay intervals rather than a 30-second interval. The delay did not increase errors, trials to first mastery or to a criterion of mastery, although time scores showed that delay subjects spent longer in the maze. Warden and Haas (1927) distinguished between learning and performance, concluding that delay has no effect on learning,

[for] even if the slowing up in speed of locomotion were clearly due to a difference in motivation arising from the interval of delay it would still be difficult to connect this in any important way with the process of fixation [p. 114].

However, Hamilton (1929), also using a maze task, found that a 1-minute delay increased both trials and time scores; also, she found that a 15-second delay in food presentation reduced the number of crossings of a charged grid for food reward. The essential difference between the two studies was that Warden and Haas (1927) detained the subjects in the goal box, as had Watson (1917); but Hamilton (1929) delayed

<sup>1</sup> This paper was prepared in conjunction with a research project supported by United States Public Health Service Research Grant No. M-6066 (A) from the National Institute of Mental Health.

the reward by using a separate delay chamber. Roberts (1930), using a problem box, was able to show that when the delay interval clearly separated the response itself from the goal, both time and trials to solution were increased. Also, Roberts noted that all subjects who received delayed reward showed an increase in errors early in learning due to a tendency to go directly to the goal-box door and that this "recency" factor initially interfered with the association of the correct instrumental response with the reinforcement. Thus Roberts, in discussing his results, identified two factors which later became important: "cues" as indirect or secondary reward and the disrupting effect of competing responses, although he did not realize the full implications of his comments at the time.

Studies by DeCamp (1920) and Clements (1928) showed that rats can learn to choose the shorter of two paths to a goal, indirectly indicating time discrimination; Sams and Tolman (1925) extended the findings by showing that rats also would choose a temporally shorter path. Wood (1933), in an experiment on chicks, demonstrated that delay of reward decreases learning efficiency, although immediate shock was given for incorrect responses. His results confirmed the findings of a similar study by Warden and Diamond (1931) on rats.

By 1933, the evidence supported these conclusions: (a) Delay of reward increased time, error, and trial scores. (b) In order for the detrimental effects of delay to occur, the indirect reward from goal-box cues had to be eliminated. (c) There was no agreement on the amount of delay necessary to interfere with learning, although responses which led to more immediate reward (in time or space) were preferred to responses which resulted in delay of reward. (d) Although the data seemed incompatible

with a simple frequency-recency postulate, there was no theoretical model to explain the data and diverse temporal gradients that had been found.

*Theoretical Treatment of the Effect of Delay: 1932-60*

*Hull's Goal Gradient.* Hull (1932) proposed his now famous goal-gradient hypothesis which stated that:

the goal reaction gets conditioned the most strongly to the stimuli preceding it, and the other reactions of the behavior sequence get conditioned to their stimuli progressively weaker as they are more remote (in time or space) from the goal reaction [pp. 25-26].

This hypothesis provided an explanation for a range of phenomena directly and indirectly related to delay, such as the choice of a shorter path to a goal (e.g., DeCamp, 1920), the elimination of errors and their order of elimination, and the rate of locomotion in a maze. Anderson (1933), by reanalyzing some previously reported data (Anderson, 1932) which had confirmed the earlier study of Sams and Tolman (1925), provided support for Hull's speed-of-locomotion hypothesis. Using a time-discrimination problem, Anderson (1933) found a direct relationship between length of delay of reinforcement and running time. An extensive study by Wolfe (1934), using both a T maze and a black-white discrimination problem, again confirmed that delay of reward produces extreme inefficiency in learning, and demonstrated how Hull's (1932) goal-gradient hypothesis could be applied to explain his delay of reinforcement gradient.

However, the effectiveness of the goal gradient as an explanation for the effects of delay did not eliminate all the inconsistencies. For example, in Wolfe's (1934) study, a delay of 1 minute seriously interfered with learning, although within 100 trials some

learning was possible with delays of 10 or 20 minutes; however, Riesen (1940) reported a discrimination study in which chimpanzees failed to learn in 600 trials with only a 4-second delay, and with pretraining he was able to extend the delay gradient to only 8 seconds.

The experimental work by Perin (1943a, 1943b) considerably clarified the factors behind the diverse durations of the temporal gradients that had been reported and added support to the view that the temporal gradient itself extended only over a short interval. Perin (1943a, 1943b) employed a Skinner box, modified so that a movement in the correct direction delivered a pellet of food after a delay interval; movement in the incorrect direction was not reinforced. In one study (1943b) the bar was removed after each correct response and in the second study (1943a), it was removed after either a correct or incorrect response. Perin's (1943a) animals showed a negatively accelerated, descending gradient with no learning of the bar-pressing response, established first by pretraining (50% level of correctness), with delays as long as 30 seconds.

These studies which followed Hull's (1932) formulation of the goal-gradient hypothesis led to a revision in his *Principles* (1943):

what was originally regarded as a single principle has turned out upon investigation to involve two fairly distinct principles: (1) the short gradient reported by Perin, which will be called the *gradient of reinforcement* . . . and (2) the more extended gradient which is presumably generated as a secondary phenomena from Perin's gradient of reinforcement acting in conjunction with the principle of secondary reinforcement. The second and more extended gradient may . . . retain the original name of the *goal gradient* [p. 143].

Thus, Hull (1943) retained the idea of a short temporal gradient of reinforcement, but treated the extended goal gradient, not as a primary principle, but

one derivable from the basic principles of the system. In his final treatment, Hull (1952) (a) derived the gradient of reinforcement from other postulates, whereas previously (Hull, 1943, p. 178) it had been treated as a postulate, and (b) distinguished between two types of delay of reinforcement (Hull, 1952, pp. 126, 132). The first case refers to a chain of responses, as occurs when reinforcement is delayed by differing runway lengths; the reinforcement was seen as decreasing in strength as it is generalized backward through the chain of responses to the beginning of the sequence, thus yielding the gradient of reinforcement within a given reaction chain (1952, p. 126, Corollary iii, A). The second case involves a single S-R connection for which reinforcement is delayed and without the subject doing anything in particular during the delay interval, the limit of learning being fixed by the duration of the stimulus trace through which the goal response can operate to reinforce the stimulus trace at the end of the interval and generalize to the beginning instrumental act (1952, p. 132, Corollary iii, B). Thus, Hull (1952) retained the notion of a short primary gradient, but derived it from other postulates of the system, and further, he identified two types of delay of reinforcement conditions; those involving a response chain and those in which there was only a single response.

*Spence's Treatment of Delay.* Perkins (1947), using a T maze, made an attempt to eliminate as many cues as possible by enclosing the maze, interchanging the two delay boxes from one side to another resulting in each delay box being followed by reward half of the time, and rotating the maze 180 degrees at the start of each day. Perkins (1947) found a limit of 120 seconds delay on the ability of the rats to solve the problem. The performance of the

experimental subjects who had the delay chambers interchanged was poorer than the performance of a control group for whom the delay chambers were not interchanged, that is, for the control subjects the same delay chamber was always followed by reward. Presumably, slight differences in the delay chambers were sufficient to provide distinctive cues for differential learning, via secondary reinforcement, of the choice task. It was the data of Perkins (1947), Riesen (1940), and Gulde (1941; cited in Grice, 1948) that led Spence (1947) to suggest that:

it would not seem unreasonable to hypothesize that there is no primary gradient of reinforcement but that all learning involving delay of primary reward results from the action of *immediate* secondary reinforcement which develops in the situation . . . . Such a hypothesis eliminates the necessity of explaining how reward seemingly acts backward over time to influence something which occurred earlier [p. 7].

Spence went on to propose that although one might eliminate secondary reinforcement from the external environment (as attempted by Perin, 1943a, 1943b; Perkins, 1947) a possible basis for differential secondary reinforcement might be proprioceptive cues existing within the animal as a result of the differential responses. Thus, Spence proposed that the concept of delay of reinforcement would be treated economically as a secondary reinforcement phenomena.

Grice (1948), in a well-controlled study, further clarified the role of secondary reinforcement in delayed-reward learning by using a discrimination problem in order to eliminate any distinctive turning response from always being associated with reward, as it is in a T maze (e.g., Perkins, 1947; Wolfe, 1934). In separate control groups he introduced either distinctive white or black goal boxes which matched the discrimination choice or obstacles which

forced the rat to make distinctive motor responses upon entering the delay chamber. With immediate reinforcement, subjects learned the discrimination in 20 trials, but with a delay of only .5 second, nearly 5 times as many trials were needed, and with a delay of 5 seconds, the number of trials increased to 580, and within the limits of the experiment (1,440 trials) no learning was possible with a 10-second delay. With the aid of the cues provided by either the distinctive goal boxes or the obstacles, there was significantly faster learning by these groups. Grice's (1948) gradient was much steeper than those of Wolfe (1934), Perin (1943a, 1943b), or Perkins (1947), and more comparable to the extremely short ones found by Riesen (1940) and Gulde (1941). Grice (1948) felt he had clearly demonstrated the effect of secondary reinforcers and felt that it was immediate secondary reinforcement which accounts for delayed reward learning; further, he suggested that the fact that any learning occurred at all under delay conditions may be due to internal cues such as sensory traces of the black-white discrimination which persist, in a decreasing fashion, for perhaps as long as 5 seconds, and that it is this trace which is contiguous with the food reward. The acceptability of Grice's (1948) speculation about stimulus traces and secondary reinforcement is a matter of theory which we will return to later, but clearly, the temporal gradient, at least in rats, appears to be very short, and distinctive cues are largely responsible for the duration of the gradient over time (i.e., its slope).

Spence (1956) revised his earlier (1947) statements on the basis of experimental studies occurring in his laboratories (Carlton, 1954; Harker, 1950; Logan, 1952; Ramond, 1954; Shilling, 1951). Spence (1956) recognized, as

did Hull (1952), that there were two types of delay of reinforcement (i.e., a response chain as in runways of varying length and a nonchaining case where the delay is introduced between the instrumental response and the reinforcement). Rather than utilizing the notion of a decreasing secondary-reinforcement effect which acts to limit the habit strength of the instrumental response, Spence suggested (a) that delay may operate instead to reduce the incentive motivation the longer the response chain over which the  $r_g$  must generalize, or in the nonchaining case, that the delay weakens the strength of the classically conditioned consummatory response; (b) that delay, of the second or nonchaining type, is analogous to experimental extinction, with the decrements in performance under delay of reinforcement resulting from essentially the same inhibitory factors that operate in experimental extinction (Spence, 1956, pp. 148-164).

Regarding the latter suggestion, the studies (Carlton, 1954; Harker, 1950; Shilling, 1951) reported in Spence (1956) show that the performance decrement under delay is contingent upon the opportunity for competing responses to occur; failure to maintain a goal orientation and the occurrence of conflicting responses thus provide a contribution to an additive inhibitory variable that reduces performance. Ramond (1954) provided some support for the additive nature of the inhibitory effect of delay by finding no interaction with drive level, as would be predicted from Hull's formulation; this finding has been replicated (Renner, 1963).

*Mowrer's Revised Two-Factor Theory.* Mowrer (1960) recently has discarded the concept of habit by proposing that all learning is conditioning, and he has provided a model that is applicable to both reward and aversive learning

under delay of reward or punishment. The literature on delay of punishment is sparse and will be treated in a separate section of this paper; however, the recent revision of two-factor theory by Mowrer (1960) may provide the common theoretical framework necessary for integrating the delay of punishment with the delay of reward literature. The revised two-factor theory seems to have its greatest importance in accounting for the effects of cue utilization and secondary reinforcement with respect to learning and extinction under delay of reinforcement.

Mowrer (1960) proposes that the affective component of the reaction made to either decremental or incremental reinforcement becomes classically conditioned to either the response produced or external stimuli. He assumes that traces of such stimuli may persist for as long as 30-45 seconds which yield four categories of temporal gradients of reinforcement. After such conditioning these stimuli become positive or negative secondary reinforcers which may "bridge" an interval between stimulation and reinforcement longer than 30-45 seconds; the conditioned stimuli, or secondary reinforcers, presumably have gradients of lesser extent and level, but are capable of establishing new habits (i.e., second-order conditioning of hope or fear). Mowrer derives support for his position by considering much of the same literature as Spence (1947) and Hull (1952), but arrives at the different conclusion that there is a temporal gradient equal to the time interval over which trace conditioning is possible. Mowrer regards the conditioned secondary reinforcers as the habit, rather than a derived reward which in turn acts as an immediate reinforcement for an S-R relationship. It is to be expected that delay of reinforcement experiments will be generated from his notions, especially

with respect to the role of both positive and negative secondary reinforcement in both delay-of-reward and delay-of-punishment learning (Mowrer, 1960, pp. 345-387).

*Theoretical Issues.* The history and the theoretical hypotheses on delay of reinforcement provide a context in which to classify and evaluate the recent experiments (primarily post-1956). Some of the questions to be considered are: What is the evidence for or against the notions of a primary gradient of reinforcement versus a derived gradient? What is the role of secondary reinforcement and cue utilization in delayed reward learning? How does delay produce its detrimental effect? If nonreward (or extinction) is the limiting case of delay of reinforcement, then, the effects of delay as a frustration, the partial reinforcement effect, and the role of competing responses should have a direct relationship to delay of reward learning. What is the similarity between delay of punishment or escape and delay of positive reinforcement with respect to the nature of the temporal gradients? What are the results of comparisons across species, especially when extended to the human level? And finally, are there some implications of these issues for more complex response patterns, such as conflict behavior where rewards and punishments are both administered, but not simultaneously? These are the more important issues in terms of which the recent literature can be examined.

#### RECENT EMPIRICAL AND THEORETICAL FINDINGS

##### *On the Nature of the Temporal Gradient*

*Primary or Derived.* The question of the existence of a "true" temporal gradient of reinforcement (e.g., Hull, 1952) versus a derived gradient (Spence, 1947, 1956) has continued to interest several

investigators. A unique experiment by Mathers (1953) attempted to impose even greater control on secondary reinforcement than Grice (1948) by eliminating any distinctive aspect of the stimulus trace of the discrimination task. Recall that Grice (1948) had speculated that it was this stimulus trace that permitted any learning at all. Mathers (1953) used a Y maze and ran his animals, on any given day, under either food or water drive, but never under both drives. On a water-deprived day, a white choice was followed by water, and on food-deprived days, black was rewarded by food, but on any given day only one reward was present; different groups experienced different delays of reinforcement. Thus, each stimulus trace was followed by reward equally often and should have acquired equal reinforcing power, therefore eliminating any differential effect occurring because one of the stimulus traces was always followed by reward. His rats learned to go to the correct side on any given day with delays of up to 20 seconds, although not with a high level of accuracy, and, as expected, subjects with a shorter delay performed consistently better than subjects with longer delay intervals. Mathers (1953) concluded that his data supported the concept of a temporal gradient.

Mowrer (1960, p. 360), by considering delayed-reward learning as trace conditioning, argues for a primary temporal gradient of 30-45 seconds duration; he considers the shorter empirical gradients (e.g., Grice, 1948; Riesen, 1940) an atypical function of the complex discrimination task used, which he regards as insensitive for studying delay of reinforcement since a sign rather than a specific response must be learned. Thus, for Mowrer, the gradient represents the decreasing effectiveness of a reinforcement as it is more remote in

time from the response-produced stimuli associated with a specific response. But, the justification is not clear for Mowrer's implicit distinction that the stimulus trace of external stimuli (as in discrimination learning) should be effective over a shorter duration than the trace of distinctive internal response-produced stimuli (as in position learning) for the conditioning of hope or fear, especially since in a discrimination problem the negative stimulus also takes on avoidance properties (Grice & Goldman, 1955) which may aid learning.

Two other investigators, one using reward with pigeons (Myers, 1958) and the other using shock and a conditioned fear response with rats (Davitz, Mason, Mowrer, & Viek, 1957) likewise interpreted their results as supporting a primary temporal gradient and refuting Spence's (1947) earlier suggestion that all learning under delay is due to secondary reinforcement. Myers (1958) concluded that although differential cues facilitate learning, they are not essential. Also, Lawrence and Hommel (1961) recently have questioned the stimulus-trace theory as put forth by Grice (1948).

The use of electrical brain stimulation would seem to offer a useful technique for studying the nature of the temporal gradient; but, apparently this method has not yet been explored. Apart from this "absolute" problem, the variables which affect the characteristics of the empirical temporal gradient (e.g., modify its slope or duration) have been investigated in their own right and have provided some interesting and theoretically useful observations.

*Secondary Reinforcement.* Frequently (e.g., Brackbill & Kappy, 1962; Grice, 1948; Renner, 1963; Spence, 1947), it has been assumed that cues present in the learning situation acquire secondary reinforcing power which may greatly

extend the duration of the temporal gradient over time. The argument is that any stimulus event which is associated with primary reinforcement becomes conditioned to it and thus provides immediate secondary reinforcement for the correct instrument response. Grice and Goldman (1955) demonstrated that in a discrimination problem the avoidance of cues associated with an incorrect response also operates to account for choice learning under delay of reward; thus, where positive and negative cues are present, both secondary reward and avoidance appear to be operating. Also there is indirect evidence that distinctive cues act as secondary reinforcers as Spence (1947) suggested; for example, Kelleher (1957), in an experiment on token learning, found that performance was inversely related to the delay in exchanging the chips for a primary reward, a delay of 1 hour leading to extinction. However, only two investigators (Jenkins, 1950; Peterson, 1956) independently demonstrated that such cues are capable of establishing a new response, and only Jenkins assessed the relative reinforcing power of such cues by obtaining a derived secondary gradient, that is, the longer the delay between the cues and the reinforcement the less effective was the cue for establishing a new response. Thus, at least in some conditions, cues become reinforcers, but the conditions under which cues are selected and utilized is not clear.

*Cue Utilization.* Several studies (e.g., Lutz & Perkins, 1960; Renner, 1963) have demonstrated that when distinctive cues are present, they facilitate the performance of subjects when reward is delayed, although these studies do not provide independent evidence that such cues actually would be capable of reinforcing a new response. Lutz and Perkins found that the length of delay made

no difference, whereas Renner suggested that the effectiveness of a cue in facilitating performance may be related to the length of the delay interval. These findings suggest that delay may be related meaningfully to the breadth of learning hypothesis (see Kausler & Trapp, 1960) as one of the factors leading to less focused behavior. In addition, the availability of positive cues may contribute to the resistance to extinction that responses learned under delay-of-reward exhibit (to be discussed later). Thus, the utilization of cues, which may then become secondary reinforcers, seems to be dependent on the delay more than on mere contiguity with reward (Lutz & Perkins, 1960; Peterson, 1956; Renner, 1963) just as partial reinforcement (Notterman, 1951) or discrimination training (Wike & McNamara, 1957b) may facilitate the reinforcing effect of a distinctive stimulus more than the mere contiguity with reward. However, further work is needed to clarify the role of cue utilization in delayed-reward learning and the relationship of cues to the breadth of learning hypothesis and secondary reinforcement.

In a related study (Holder, Marx, Holder, & Collier, 1957), the effects of delay (interpreted as frustration) was conditioned to a conditioned stimulus (CS); which in turn, prompted faster running in a second runway because of the increased motivation resulting from the conditioning of the delay-produced frustration drive to the CS. This study demonstrated that delay can also result in an external stimulus taking on the power of a negative secondary reinforcer. The second implication of this study—that is, the motivational aspects of lack of an expected reinforcer—will be discussed later; what remains to be clarified is whether cues brought into operation by or during the delay are facilitating performance by acting as

positive secondary reinforcers which span the time interval, or if such cues take on negative secondary reinforcing power and facilitate postdelay performance through an increase in the motivation. The details of cue utilization under delay of reinforcement remains to be specified, but delay versus goal-associated cues may be a necessary distinction.

*Interactions.* The runway studies of Ramond (1954) and Renner (1963) have demonstrated that drive level and delay combine additively, with an increase in drive level elevating the gradient. Apparently, the Drive Level  $\times$  Delay interaction has not been studied in a choice situation.

Logan (1960, p. 58) reports that delay and amount of reward combine in additive fashion to determine response speed; again, data are available only for runway performance and not for choice behavior. However, as Logan (1960) cautions,

[the] units-of-measurement question is of special significance in interpreting interactions . . . because on only the assumption that response speed is a linear estimate of response tendency can the obtained additive interactions in performance be taken as indicating additive interactions among the relevant variables [p. 60].

It is clear that the empirical temporal gradient of reinforcement is a negatively accelerated decreasing function, but the exact nature of the gradient is still in question, and how it is modified by other variables in a variety of problems is yet to be specified. More work will be required in order to establish the manner by which delay of reinforcement combines with other variables to control learning and performance.

*Other Observations.* Ferster (1953) and Dews (1960) found that a pecking response could be maintained with a delay that ordinarily would have produced extinction (e.g., 60 seconds), by first establishing the response with short delays and gradually increasing the delay inter-

vals. Consistent with this is the report by Lawrence and Hommel (1961) that delays in a Grice box can be extended to 60 seconds without any decrease in accuracy after a discrimination had been learned. Thus, the experience of the organism with a specific response is a relevant variable in the ability to function with delayed reinforcement, at least with respect to maintaining an established response (Dews, 1960), although there is apparently no evidence on the possible existence of a learning how to learn-under-delay of reinforcement phenomena. It seems reasonable to expect that experience with delay of reinforcement may be an important factor in the general ability of an organism to function under delay of reinforcement.

Levin and Forgays (1960) have demonstrated that sensory change can act as a reinforcer in a maze, and that delay in the sensory change leads to a delay-of-reinforcement gradient similar to those resulting from the use of traditional primary reinforcers.

One of the most interesting and systematic exploration of delay of reinforcement is that of Logan (1960; Logan, Beier, & Ellis, 1955) who has studied both varied and correlated delay of reinforcement. He finds that when reinforcement is varied between two equally likely values (e.g., 1 and 9 seconds) that it will lead to faster performance than at a constant delay at the mean of the two values (e.g., 5 seconds) and approaches the performance of a group with a constant delay at the shorter of the two values (e.g., 1 second) (pp. 70-71, 73-74). Peterson (1956) failed to confirm this last aspect of variable delay as have Wike and Kintsch (1959) with partial delay; however, Pubols (1962) has confirmed the former conclusion for choice behavior, that is, subjects will show a preference for a varied reward over a constant delay at the mean of the two

values; this finding follows from the assumption of a negatively accelerated decreasing gradient, which Logan (1960, p. 90) derives from the assumption that incentive is related to the  $r_g$  which becomes conditioned to the traces of cues (stimuli) present when the instrumental response was completed.

With respect to correlated delay of reinforcement (Bowrer, 1961; Logan, 1960), a subject can be trained to run more slowly by delaying reward when his speed exceeds a given level. Davenport (1962) has extended this type of work by correlating the amount of reward with delay (e.g., large reward with long delay versus small reward with short delay) producing equal reinforcement value contours, that is, there are critical delay values at which the two response strengths for two different reward magnitudes are equal. A unit increase in delay can be balanced by a log-unit increase in magnitude.

Studies which pit one major variable against another, or in Logan's (1960, p. 69) terms, negatively correlated reinforcement, are of theoretical and empirical importance, for the information provided about the relative effects on performance of manipulations of such major variables as delay, deprivation, and amount of reward, should lead to a clarification of their corresponding intervening variables and of behavior theory in general. It is perhaps too early to draw any conclusions, but it seems apparent that delay of reinforcement is one of the relevant variables that will contribute to this approach, and a logical extension of Logan's work is to study choice behavior (e.g., Davenport, 1962; Pubols, 1962).

### *Delay of Reinforcement and Extinction*

Delay of reward is associated with a weaker response strength as long as an acquisition measure is used; however,

when an extinction measure is utilized, at least under some conditions, the tentative statement can be made that delay of reward is associated with a stronger response strength. The initial observation (Crum, Brown, & Bitterman, 1951) of the increased resistance to extinction following partial delay of reinforcement has stimulated a series of experiments and raised enough problems to make any simple explanations of the effects of delay of reinforcement unlikely and has linked delay of reward to theories of partial reinforcement.

The study by Crum et al. (1951) has been confirmed in a reversal task (Kintsch & Wike, 1957) and in runway studies (e.g., Logan, Beier, & Kincaid, 1956; Peterson, 1956; Scott & Wike, 1956; Wike & McNamara, 1957a). Two failures of confirmation are reported where partial delay of reinforcement did not lead to a greater resistance to extinction, the first in a T maze (Wike & Remple, 1959) and the second in a runway (Wike & Kintsch, 1959); however, both of these studies used less than the most effective delay interval. The latter study did not delay the subjects during extinction but still obtained results in the expected direction, and the former study ran an insufficient number of trials (18 free and 18 forced) as indicated by the fact that one delay group was performing well below chance and all groups were considerably below asymptotic level. There is sufficient evidence to conclude that at least under conditions where the delay interval is relatively long (e.g., 30 seconds) and is present on both acquisition and extinction trials, partial or variable delay of reinforcement increases resistance to extinction, although the exact role of such complicating factors as patterning effects, trial distribution, cues, secondary reinforcement, and delay-produced frustration are far from being understood.

But, even more fundamentally, it is not clear if the crucial factor in producing this resistance to extinction is the delay per se, or if it is the variability, that is, the fact that the delay is present on only part of the trials. The few studies which used constant delay have yielded inconsistent results. Fehrer (1956) using a runway, and Pubols (1958) using a discrimination reversal problem, found that constant delay of reinforcement alone is sufficient to increase resistance to extinction; however, Logan et al. (1956; Logan, 1960) have obtained negative results and Renner (1963) inconclusive results. Thus, in the few studies employing constant delay of reinforcement, it is not clear if delay alone is sufficient to produce resistance to extinction. More experimental work is needed to first demonstrate conclusively if delay of reinforcement during acquisition is sufficient to produce resistance to extinction, or if it is the partial or variable aspect that is necessary, and second, to explain what the crucial factors are in producing the apparent contradictions reported thus far.

The diverse experimental findings on extinction have made it, as a general phenomena, difficult for any general theory to handle (see Kimble, 1961); and just as partial reinforcement has provided a dilemma, so possibly may delay of reinforcement present some unique problems. Logan (1960) considers the problems of delay and partial reinforcement as essentially identical, for partial reinforcement is the extreme case of simultaneously varied delay and amount of reward with a perfect positive correlation. "Partial reinforcement as the limiting case of increasing the range of variation in reward (e.g., delay and amount) produces the greatest resistance to extinction [Logan, 1960, p. 195]." Thus, the greater the variation in partial (or variable) delay, the more

identical it is to the limiting case of partial reinforcement and the greater the resistance to extinction; the data essentially support this prediction (Kintsch & Wike, 1957; Logan, 1960; Logan et al., 1956) and the prediction is consistent with the literature on partial delay (e.g., Crum et al., 1951; Scott & Wike, 1956; Wike & McNamara, 1957a) in which a variable 0-30 seconds delay pattern is necessary in order to produce the greater resistance to extinction. Recall, 30 seconds has been considered as the maximum limit of the temporal gradient (Hull, 1952). If Logan's (1960) position can be supported, then the delay problem has been translated into the old problem of partial reinforcement or vice versa. However, two comments are relevant:

First, additional study of constant delay of reinforcement is needed, for if this too retards extinction (Pubols, 1958), then there is a strong case for an additional paradox where by delay itself retards the growth of a habit but results in slower extinction; if this is so then delay must be treated as a unique case rather than as a mere variation of partial reinforcement.

Second, the extinction studies on partial delay of reinforcement were primarily concerned with furthering an understanding of partial reinforcement and extinction. Several studies were initiated to test the Hull-Sheffield generalized-decrement hypothesis (e.g., Crum et al., 1951), or interpreted in terms of the Mowrer-Jones discrimination hypothesis (e.g., Fehrer, 1956), or in terms of extinction of the occurrence of  $r_g$  to the cues of nonreinforcement (Logan, 1960), or in terms of frustration theory (e.g., Renner, 1963). Most of the literature concerning the effects of delay on extinction has been interpreted in terms of one of these theories, and the study of delay of reinforcement may contribute

to a further testing and modification of these theories.

Thus, delay is relevant to a number of theoretical issues, for example, delay may cause frustration which in turn may effect drive level; it may introduce competing responses or produce conflict through conditioned avoidance, and such effects, and/or the delay itself, may influence the degree to which cues are utilized and the positive (Grice, 1948) or negative (Grice & Goldman, 1955) secondary reinforcing power of such cues; this account by no means exhausts the list of potential complexities, but does, perhaps, emphasize the point that any comprehensive treatment of delay has not yet been achieved. Before turning to a consideration of additional theoretical problems to which the effects of delay are relevant, it may be well to examine first in greater detail the consistencies and contradictions between constant and partial (or variable) delay of reinforcement.

The major inconsistency is the failure, on some occasions, of constant delay to yield increased resistance to extinction; however, where greater resistance has not been found the subjects have been removed at the end of the runway without the customary detention (Logan, 1960, pp. 185-186; Logan et al., 1956) in the delay chamber, or delayed in the goal box rather than in a delay chamber (Wike & Kintsch, 1959). Positive results have been reported (Pubols, 1958) when such variations did not occur during extinction. Logan (1960, p. 185) in qualifying his conclusion that constant delay does not increase resistance to extinction, noted that resistance to extinction may be a function of the delay interval (detention) during acquisition and detention during extinction. Detention time on nonreinforced trials (or delayed trials) may be highly relevant, as indirectly supported by Pubols (1962) who

found different rates of learning to be a function of time spent in the delay chamber on nonreinforced trials, or as directly supported by Peterson (1956) who found resistance to extinction a direct function of time spent in the delay chamber, by Kintsch and Wike (1957) who found resistance to extinction was increased by a greater length in partial delay, and by Wike and McNamara (1957a) who found partial support for the hypothesis that resistance to extinction is related to the percentage of delayed trials. Thus, constant delay as well as partial delay may produce increased resistance to extinction proportional to the length of delay, but, the effect may be obscured for constant delay by the fact that partial- (or variable-) delay groups may reach a terminal acquisition level of performance similar to the immediate group, whereas the constant-delay group has a consistently lower terminal performance level, and although their rate of extinction (slope of curve) may be slower than the immediate group, no absolute differences are found (Renner, 1963; Wike & Kintsch, 1959). In contrast, groups trained under variable or partial delay may start extinction at essentially the same initial point as the immediate group and clearly diverge from the immediate group during extinction (Logan et al., 1956). To date, no study has provided suitable or complete comparisons between immediate reinforcement, partial (or varied), and constant delay, but there is some evidence that when immediate and constant-delay groups are similar, that constant delay increases resistance to extinction (e.g., Pubols, 1958; his subjects reached a similar acquisition criterion, although the amount of overlearning was not identical).

#### *Theoretical Implications and Problems for Research*

As was implied in the last section,

there are probably insufficient empirical data to give a systematic account of delay of reinforcement that will explain all of the acquisition and extinction aspects of the phenomena; but there are several theoretical positions that are related to performance under delay and which provide numerous, albeit fragmented, accounts for the effects of delay of reinforcement.

*Frustration Theory.* Frustration theory (Amsel, 1958; Amsel & Roussel, 1952) is able to provide a rather complete theoretical account of the effects of delay. Presumably, the introduction of a delay should be frustrating and produce conditioned avoidance and cause a decrement in performance early but not later in learning, as was observed by Renner (1963), because late in learning the frustration should become conditioned to running; however, during extinction, when the immediate-reward subjects encounter nonreward and frustration for the first time, differential rates of extinction should be observed. Thus, resistance to extinction under delay may be dependent upon prior practice in performing the instrumental response in the presence of delay (frustration) cues. For example, Davenport (1962) observed an inhibitory function of delay (presumably a frustration) which increases during the early learning trials; also Fehrer (1956) noted an absence of any distinct emotional response during extinction by subjects that had experienced delay, but in other animals who had not experienced delay, the emotional reaction was present and the subjects extinguished faster.

Holder et al. (1957) found that an intramaze delay decreased runway speed before the delay interval and increased speed in the segment after the delay. Presumably, the frustration and conditioned avoidance (Amsel, 1958) depressed initial runway performance, but the increased drive resulting from the

frustration increased subsequent performance. However, this confirmation of frustration theory by Holder et al. (1957) was not replicated in a similar study by Wist (1962).

*Delay and Cue Utilization.* As was noted previously, delay of reinforcement apparently can act to facilitate the use of cues. Perhaps the important effect of delay (detention) with respect to increased resistance to extinction is to increase the breadth of learning (Esterbrook, 1958; Kausler & Trapp, 1960) and facilitate the possible role that cues may serve. Distinctive delay-associated cues which initially may become conditioned to the negative effects of frustration later may be counter conditioned so that greater tolerance for responding in the presence of frustration-associated cues is acquired, thus producing an initial performance decrement but later contributing to resistance to extinction. This interpretation would mean that differential resistance to extinction between delay and nondelay groups would be a joint function of: greater tolerance for nonreinforcement for the delay groups and frustration and avoidance by the nondelay group who encounter nonreinforcement for the first time during extinction. Whereas, distinctive goal-associated cues, that is, presumably time-bridging secondary reinforcers, may facilitate or make possible extended delayed-reward learning and, as secondary reinforcers present during extinction trials, increase the resistance of the response to extinction, although again the delay interval may be necessary to activate the cues.

The important role of cues is well illustrated by the Perkins (1947) study in which interchanging of presumably identical delay boxes significantly interfered with the solution of a maze problem under conditions of delayed reinforcement. Thus, detention time, cue utilization, and the possible reinforcing

(Peterson, 1956) or inhibiting effects (Grice & Goldman, 1955) of such cues may be expected to provide some complex and puzzling interactions.

*Competing Responses.* Although an interference theory in its traditional sense is probably an incomplete explanation there is ample evidence that opportunity for competing responses (Spence, 1956, pp. 163-164) exists under delay conditions. In fact, one study (Pubols, 1958) provided an analysis of the acquisition and extinction data in terms of competing responses. The possibility exists that resistance to extinction may vary with the number of competing responses that have been extinguished over acquisition trials and must be readapted in order to break down the habit during extinction. Thus, some systematic variation in the opportunity for such responses may be correlated with slower acquisition and greater resistance to extinction. Unfortunately, there is no study of the relationship between a range of possible competing responses (as one type of interference) and resistance to extinction. It does not seem likely that a response competition theory is sufficient, but, as was noted in the discussion of frustration, which is also a brand of interference theory, the empirical data (e.g., Davenport, 1962; Renner, 1963) warrant the conclusion that delay does have inhibitory functions that accumulate with trials, especially early in learning (recall the early observations of Roberts, 1930).

The conclusions reached to this point may be summarized as: (a) Constant delay of reinforcement will retard acquisition. (b) Partial or variable delay may or may not limit acquisition-response strength depending on the experimental conditions, that is, number of trials, percentage of delayed reinforcements, and length of delay intervals. (c) Constant delay may retard resistance to extinction, providing: cues are pres-

ent to be "activated" by delay; such cues act as secondary reinforcement for the response and/or making the response to frustration associated cues is learned; the detention during acquisition and extinction is similar so as not to aid the discrimination between acquisition and extinction conditions; that appropriate comparisons can be made, that is, delay and immediate groups have reached a similar criterion so that there is sufficient opportunity for the extinction curves to diverge at some point above no performance. (d) Variable delay will lead to increased resistance to extinction the longer the maximum delay and possibly the higher the percentage of delayed trials, and is essentially the same phenomenon as partial reinforcement. (e) The effects of delay combine additively with such other variables as drive and amount of reinforcement.

#### ADDITIONAL TEMPORAL ASPECTS OF REINFORCEMENT

##### *Delay of Punishment*

The effectiveness of delayed punishment versus delayed reward has not been explored. Kimble (1961) reports only three studies (Bevan & Dukes, 1955; Kamin, 1959; Warden & Diamond, 1931) on the delay of punishment gradient and comments that it "would be interesting to know whether this gradient is approximately the same for the situation in which the punished response was established with a positive reinforcer [p. 155]." Although Kimble's question cannot be answered, there are a few general statements that can be made.<sup>2</sup> Two of the three studies obtained a temporal gradient similar to ones obtained for

food reinforcement. Kamin (1959) obtained a temporal gradient by punishing an avoidance response after varying intervals of delay; the inhibitory effects of the punishment declined as a function of delay and reached an asymptotic level at 40 seconds. The slope of this temporal gradient of inhibition was more or less comparable with reward established gradients, but not with the flat inhibitory gradient reported by Bevan and Dukes (1955), which was produced by punishing the incorrect choice in a Y maze. However, Bevan and Dukes' (1955) study and the early study by Warden and Diamond (1931) involved the use of food reward; thus, no uncomplicated direct comparison between delay-of-reward and delay-of-punishment gradients can be made. The only apparent "pure" delay-of-punishment gradient in which the punishment is used to establish an instrumental response is the unpublished study of Butler (1959) which confirms the earlier studies by yielding gradients similar to those obtained by reward.

Additional evidence on the nature of the temporal gradient of punishment can be found in a study by Azrin (1956) or in the pair of experiments by Kamin (1957a, 1957b); these later studies demonstrated a gradient of delay of secondary reward (delay in CS termination) on avoidance learning.

The apparent parallel between delay of reward and delay of punishment is highlighted by the Crum et al. (1951) study which showed that varied delay of punishment leads to greater resistance to extinction than immediate punishment. Thus far there have been no systematic studies on extinction of instrumental responses established under delay of punishment.

Sidman (1954) indirectly demonstrated a delay-of-punishment gradient; his subject could postpone the occurrence of a brief shock by either a constant or a variable time interval, depending on which of two bars was pressed.

<sup>2</sup> Delay of reward and delay of punishment experiments which use the same apparatus and require an identical instrumental position response currently are being conducted. The acquisition and extinction of these responses, established under varying drive and cue conditions, should yield further information on the delay of reward and delay of punishment gradients.

Small time differences between the bars had an effect on the response when the delay interval was short, but when the delay is long no learning can occur without a larger time difference between the bars.

There are many unknowns with respect to the temporal gradient of punishment. The few studies available suggest a similarity with reward data, but it is doubtful if, as empirical data accumulate, it will be more than just a superficial similarity. For example, there are indications in Bevan and Dukes' (1955) experiment that shock may produce anxiety that disrupts the gradient. A further confusing note is the finding (Knapp, Kause, & Perkins, 1959) when cues are very distinct and a delay in food reward is constant regardless of choice, a subject will prefer immediate shock and then a delay interval in preference to delayed shock (delay of 30 seconds and then shock) because the distinct differential delay stimuli acquire conditioned fear arousal and the immediate occurrence of the shock permits preparatory responses which may decrease its intensity. This interpretation of their data requires that distinctive differential stimuli are present; by way of contrast, Butler's (1959) study offered subjects a choice between immediate or delayed punishment, and the preference was clearly for delayed punishment, but his situation had few if any distinctive cues, whereas the Knapp et al. (1959) study had many cues, and in addition at the end of the delay, provided food reward in order to elicit the choice response in the first place. This procedure limits the range of shock that can be used.

#### *Delay of Reinforcement with Human Subjects*

Delay of reinforcement largely has been confined to the study of animals, especially rats, and it is only recently that a number of studies with human

subjects have appeared. The studies are only roughly comparable in that many of them use delayed feedback of knowledge of results on the learning of some rote skill rather than delay of a primary reward. A further problem is the role of verbal cues which are out of the experimenter's control, but perhaps very influential. The important comparative question is whether the basic generalizations available from animal experiments also apply to human learning situations. Specifically, does delay retard acquisition, does it influence retention or extinction, and what role do "cues" play in changing the acquisition or extinction (forgetting) of responses learned under delay of reinforcement?

The majority of studies have been concerned with the effect of a delay of knowledge of results on performance of a simple motor task and with one exception (Greenspoon & Foreman, 1956) have yielded negative results, that is, delay of knowledge of results does not interfere with performance (Bilodeau & Bilodeau, 1958; Bilodeau & Ryan, 1960; Denny, Allard, Hall, & Rokeach, 1960; McGuigan, 1959; Noble & Alcock, 1958; Saltzman, Kanfer, & Greenspoon, 1955); one additional study had inconclusive results (Ryan & Bilodeau, 1962). These studies have used delays that ranged up to 7 days (Bilodeau & Bilodeau, 1958; Ryan & Bilodeau, 1962) although the typical range was from 0 to 30 seconds. On simple motor tasks, delay of knowledge of results does not cause a performance decrement, and the studies by Bilodeau and Bilodeau (1958) and Denny et al. (1960) demonstrate that the crucial delay is the time between responses (intertrial interval) and not the time between response and feedback; that is, only long intervals between responses are damaging to performance.

In experiments where some verbal skill or memory is involved, the results are contradictory. Saltzman (1951) and Bourne (1957) reported performance

decrements associated with increasing delay in knowledge of results; however, Brackbill and Kappy (1962) using a serial-learning task reported delay had no effect (the delay intervals in these studies were all comparable). The essential difference was attributed by Brackbill and Kappy (1962) to the extent to which response-produced cues operate in the situation. They reasoned that delay has no effect when cues are present, but if the situation is such that the subject is responding without cues in a random trial-and-error fashion, then delay of reward (or knowledge of results) will decrease performance; when the subject is able to link the consequences with the response, delay has no effect. Brackbill and Kappy's (1962) hypothesis may be evaluated by considering two other lines of evidence.

First, in a series of experiments with children in a simple response or discrimination problem it was found that, similar to rat studies, children would prefer immediate reward and respond faster to immediate reward (Lipsitt & Castaneda, 1958; Lipsitt, Castaneda, & Kemble, 1959) and that delay of reinforcement would impede acquisition of a discrimination response (Hockman & Lipsitt, 1961). However, when subjects were instructed to pay attention to the goal (goal-directed instruction), delay of reinforcement on a discrimination task did not produce an effect (Erickson & Lipsitt, 1960). The implication, following Brackbill and Kappy's (1962) hypothesis, is that the delay interval has no effect when the subject can mediate the temporal interval by some cue, for example, a verbalization (Brackbill & Kappy, 1962), a motor response (e.g., Bilodeau & Bilodeau, 1958), or a goal orientation (Erickson & Lipsitt, 1960). In terms of the rat literature of the previous section, cues or secondary reinforcers control the slope of the empirical temporal gradient of reinforcement.

Second, although the argument of

Brackbill and Kappy (1962) appears to make sense for the typically short (0-30 seconds) delay intervals that have been used, the long 7-day delay interval reported by Bilodeau and Bilodeau (1958) would seem to be too long an interval to be bridged without decay by the mere presence of a cue. However, the response used (drawing a 3-inch line; information on accuracy was delayed 7 days) was based on a unit, the inch, familiar to the experience of all subjects; when unfamiliar units are used (Denny et al., 1960) the task is made more difficult and accuracy over time is reduced. The extended interval appears to be an artifact of practice with the response. Again the crucial delay in accounting for performance decrement was the delay between successive responses and not delay in knowledge of results. When the crucial intertrial interval is taken into account, delay has no effect on performance, at least on the simple motor responses studied. Brackbill and Kappy (1962) argue that the effect of delay depends on the absence of cues. This is, in some ways, a theoretical return to the Spence (1947) theory and is perhaps suitable for the limited human data available.

However, Brackbill and Kappy's (1962) study deserves attention in another respect, for they also measured retention, which may be considered in some ways analogous to extinction measures. Subjects were tested for recognition and for savings after first learning a serial list under various delays of reinforcement. The savings method presumably measured the "persistence" of the responses learned under varying delays. Delay did not interfere with acquisition but did facilitate retention measures the next day (but not 8 days later). Thus a constant delay during acquisition was associated with greater resistance to forgetting. Although no traditional extinction measure was taken, Brackbill and Kappy (1962) compared

their results to the animal work on extinction. They concluded: (a) Rats do not use cues and thus delay retards acquisition. (b) With rats, variable delay, but not constant delay, increases resistance to extinction; however, partial or variable delay is a form of partial reinforcement and belongs with that literature. By implication, their argument is that cues eliminate the performance decrement produced by delay and the utilization of cues during acquisition (as in their experiment) is what increases retention or resistance to extinction.

Their analysis requires comment. In many respects it is similar to the conclusion presented earlier based on the rat data; that is, delay (detention) facilitates cue utilization and cues may influence extinction since resistance to extinction appears to be related to detention time and opportunity to utilize the cues. The problem is probably not as simple as Brackbill and Kappy (1962) imply, for as has been indicated, cues may have several effects besides producing resistance to extinction or slower forgetting and may even elicit competing responses (Rieber, 1961). Further, contrary to their analysis, response-produced cues are available to rats which affects the slope of the temporal gradient as was demonstrated by Grice (1948). The effect of such response-produced cues (or external cues) on the extinction of a response learned under delay has yet to be systematically demonstrated in humans or rats. Finally, it is too early to conclude as Brackbill and Kappy (1962) do that with animal subjects delay has no effect on resistance to extinction (e.g., Fehrer, 1956; Pubols, 1958).

### *Delay of Reinforcement and Abnormal Behavior*

Mowrer and Ullman (1945) were the first to give systematic attention to the relevance of delay of reinforcement to abnormal and neurotic behavior. They

noted that neurotics often persist in making responses which lead to punishment. The resistance of symptoms to extinction as a result of nonreward or punishment led these authors to label the situation the "neurotic paradox" which aptly characterizes symptoms from the standpoint of reinforcement theory. Mowrer and Ullman (1945) went on to demonstrate how this paradoxical element could be explained by the temporal gradient of reinforcement by assuming that a symptom leads to some immediate reinforcement (e.g., anxiety reduction or secondary gain) and punishment after a delay (e.g., isolation, criticism, etc.). If the punishment were delayed, even for a short interval, the sharp decline of the temporal gradient could lead to a situation where the net difference could be reinforcing to the neurotic response, even though in absolute terms the punishment might have been severe enough to inhibit the response if it had been administered immediately. Mowrer and Ullman (1945) presented a demonstration rat study to support their theory. Thus, the paradoxical element was eliminated and made consistent with reinforcement theory. The only apparent experimental test with human subjects of the Mowrer-Ullman (1945) theory led to inconclusive results (Freihage, 1957).

Dollard and Miller (1950) utilized the gradient to explain why punished responses that were preceded by some reward would persist, but felt the reverse was still paradoxical to reinforcement theory. They wrote "where the increase in misery is delayed the principle of reinforcement can easily handle the problem; where an apparent increase in misery is immediate a clear dilemma is posed [p. 187]." Renner (1961) has tentatively indicated how the Mowrer-Ullman (1945) model could be extended to cover a wider range of abnormal behavior including the case of immediate punishment and delayed reward.

### *Preference for Delay of Reinforcement as a Personality Variable*

A recent series of experiments has studied the ability of children to tolerate a delay of reinforcement. The procedure provides the child with a choice between an immediate-smaller reward versus a delayed-larger reward, the implication being that the ability to postpone gratification is a sign of maturity and reality contact (Mischel & Metzner, 1962); whereas preference for immediate-smaller gratification may indicate a more self-indulgent, impulsive, and immature personality.

Mischel in a series of experiments (Mischel, 1958, 1961a, 1961b, 1961c; Mischel & Metzner, 1962) has found preference for larger-delayed reinforcement to be associated with such variables as father in the home (1958), age (1958, 1962), social responsibility (1961c, 1961a), accuracy in time concepts and time perspective (1961c, 1962), need achievement (1961a), and intelligence (1962). Consistent results on time perspective and intelligence, but not age, were also found by Melikian (1959).

The central concept is that the ability to delay gratification is acquired with age as a function of experience and that this capacity has specific personality correlates that are essentially equivalent to the level of social maturity. Two aspects of Mischel's work are relevant to this paper: (a) that the ability to accept delay of reward may be an acquired ability that changes with age and with experience with reinforcing agents (also, Mahrer, 1956), and (b) that delay of gratification may be related to abnormal behavior through acquired personality mechanisms that render delayed consequences ineffective for the modification of behavior. From this point of view it is important to know not only how delay may be related to a learning paradigm of neurotic behavior (e.g., Mowrer & Ullman, 1945), but how particular learn-

ing experiences or expectancies in the Rotter (1954) sense can develop which influence the nature of the temporal gradient of reinforcement. The last comment indicates still another reason why data on a learning-how-to-learn-under-delay process (discussed earlier) would be useful, although none are currently available.

### *Concluding Remarks*

The topics covered have been too numerous to attempt a complete summary, but several features are worth restating. It appears that delay during acquisition limits the strength of a response as long as some learning measure is used; however, when resistance to extinction is used as the measure then the effect appears to be similar to the partial reinforcement effect. The relationship between constant delay, partial (or variable) delay, and partial reinforcement is not clear; but the role of cues, both external and response-produced, appears to be important and complex. It was suggested that delay facilitates the utilization of cues but that the cues may have varied effects, depending upon the experimental conditions. The data on human delay of reward learning tend to agree in essence with the data from animal studies. Despite the long history of delay of reinforcement the vast majority of work has been with reward and there is little data on punishment; the data that exist suggests similar temporal gradients of punishment and reward. More experimental work is needed in order to clarify the effects of delay of reinforcement.

An interesting, and potentially significant, application of the delay-of-reinforcement literature to abnormal behavior and personality characteristics has been illustrated by several studies, but a realization of this application remains to be done.

REFERENCES

- AMSEL, A. Drive properties to the anticipation of frustration. *Psychol. Bull.*, 1958, 55, 102-119.
- AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effects on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, 43, 363-368.
- ANDERSON, A. C. Time discrimination in the white rat. *J. comp. Psychol.*, 1932, 13, 27-55.
- ANDERSON, A. C. Runway time and the goal gradient. *J. exp. Psychol.*, 1933, 16, 423-428.
- AZRIN, N. Some effects of two intermittent schedules of immediate and non-immediate punishment. *J. Psychol.*, 1956, 42, 3-21.
- BEVAN, W., & DUKES, W. F. The effectiveness of delayed punishment on learning performance when preceded by premonitory cues. *Psychol. Rep.*, 1955, 1, 444-448.
- BILODEAU, E. A., & BILODEAU, I. MCD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, 55, 603-612.
- BILODEAU, E. A., & RYAN, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *J. exp. Psychol.*, 1960, 59, 414-419.
- BOURNE, L. E. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, 54, 201-207.
- BOWLER, G. H. Correlated delay of reinforcement. *J. comp. physiol. Psychol.*, 1961, 54, 196-203.
- BRACKBILL, YVONNE, & KAPPY, M. S. Delay of reinforcement and retention. *J. comp. physiol. Psychol.*, 1962, 55, 14-18.
- BUTLER, D. Two-choice behavior as a function of the delay between response and punishment. Unpublished doctoral dissertation, Northwestern University, 1959. (*Dissert. Abstr.*, 1959, 19, 2662)
- CARLTON, P. L. Response strength as a function of delay of reward and physical confinement. Unpublished master's thesis, University of Iowa, 1954.
- CLEMENTS, F. E. Effect of time on distance discrimination in the albino rat. *J. comp. Psychol.*, 1928, 8, 317-324.
- CRUM, JANET, BROWN, W. L., & BITTERMAN, M. E. The effect of partial and delayed reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1951, 64, 228-237.
- DAVENPORT, J. The interaction of magnitude and delay of reinforcement in spatial discrimination. *J. comp. physiol. Psychol.*, 1962, 55, 267-273.
- DAVITZ, J. R., MASON, D. J., MOWRER, O. H., & VIEK, P. Conditioning of fear: A function of the delay of reinforcement. *Amer. J. Psychol.*, 1957, 70, 69-74.
- DECAMP, J. E. Relative distance as a factor in the white rat's selection of a path. *Psychobiology*, 1920, 2, 245-254.
- DENNY, M. R., ALLARD, M., HALL, E., & ROKEACH, M. Supplementary report: Delay of knowledge of results, knowledge of the task, and intertrial interval. *J. exp. Psychol.*, 1960, 60, 327.
- DEWS, P. Free-operant behavior under conditions of delayed reinforcement: I. CRF-type schedules. *J. exp. Behav.*, 1960, 3, 221-234.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1958, 66, 183-201.
- ERICKSON, MARILEY, & LIPSITT, L. P. Effects of delayed reward on simultaneous and successive discrimination learning in children. *J. comp. physiol. Psychol.*, 1960, 53, 256-260.
- FEHRER, E. Effects of amount of reinforcement and of pre- and postreinforcement delays on learning and extinction. *J. exp. Psychol.*, 1956, 52, 167-176.
- FESTER, C. B. Sustained behavior under delayed reinforcement. *J. exp. Psychol.*, 1953, 45, 218-224.
- FREIHAGE, M. An experimental analogue of the "neurotic paradox" in normal human subjects. Washington, D. C.: Catholic Univ. America Press, 1957.
- GREENSPOON, J., & FOREMAN, S. Effect of delay of knowledge of results on learning a motor task. *J. exp. Psychol.*, 1956, 51, 226-228.
- GRICE, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, 38, 1-16.
- GRICE, G. R., & GOLDMAN, H. M. Generalized extinction and secondary reinforcement in visual discrimination learning with delayed reward. *J. exp. Psychol.*, 1955, 50, 197-200.
- GULDE, C. J. The effects of delayed reward on the learning of a white-black discrimination by the albino rat. Unpublished master's thesis, University of Iowa, 1941.
- HAMILTON, E. LOUISE. The effects of delayed incentive on the hunger drive in the white rat. *Genet. Psychol. Monogr.* 1929, 5(2), 131-209.
- HARKER, G. S. An experimental investigation of the effects of changes in the delay of reinforcement upon level of performance of an instrumental response. Unpublished doctoral dissertation, University of Iowa, 1950.
- HOCKMAN, C. H., & LIPSITT, L. P. Delay-of-

- reward gradients in discrimination learning with children for two levels of difficulty. *J. comp. physiol. Psychol.*, 1961, 54, 24-27.
- HOLDER, W., MARX, M., HOLDER, ELAINE, & COLLIER, G. Response strength as a function of delay of reward in a runway. *J. exp. Psychol.*, 1957, 53, 316-323.
- HULL, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, 39, 25-43.
- HULL, C. L. *Principles of behavior: An introduction to behavior theory*. New York: Appleton-Century-Crofts, 1943.
- HULL, C. L. *A behavior system*. New Haven: Yale Univ. Press, 1952.
- HUNTER, W. S. Delayed reactions in animals and children. *Behav. Monogr.*, 1913, No. 2.
- JENKINS, W. O. A temporal gradient of derived reinforcement. *Amer. J. Psychol.*, 1950, 63, 237-243.
- KAMIN, L. J. The gradient of delay of secondary reward in avoidance learning. *J. comp. physiol. Psychol.*, 1957, 50, 445-449. (a)
- KAMIN, L. J. The gradient of delay of secondary reward in avoidance learning tested on avoidance trials only. *J. comp. physiol. Psychol.*, 1957, 50, 450-456. (b)
- KAMIN, L. J. The delay of punishment gradient. *J. comp. physiol. Psychol.*, 1959, 52, 432-437.
- KAUSLER, D. H., & TRAPP, E. P. Motivation and cue utilization in intentional and incidental learning. *Psychol. Rev.*, 1960, 67, 373-379.
- KELLEHER, R. T. Conditioned reinforcement in chimpanzees. *J. comp. physiol. Psychol.*, 1957, 50, 571-575.
- KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning*. (2nd ed.) New York: Appleton-Century-Crofts, 1961.
- KINTSCH, W., & WIKE, E. L. Habit reversal as a function of length of partial delay of reinforcement. *Psychol. Rep.*, 1957, 3, 11-14.
- KNAPP, R. K., KAUSE, R. H., & PERKINS, C. C. Immediate versus delayed shock in T-maze performance. *J. exp. Psychol.*, 1959, 58, 357-362.
- LAWRENCE, D., & HOMMEL, L. The influence of differential goal boxes on discrimination learning involving delay of reinforcement. *J. comp. physiol. Psychol.*, 1961, 54, 552-555.
- LEVIN, H., & FORGAYS, D. G. Sensory change as immediate and delayed reinforcement for maze learning. *J. comp. physiol. Psychol.*, 1960, 53, 194-196.
- LIPSITT, L. P., & CASTANEDA, A. Effects of delayed reward on choice behavior and response speeds in children. *J. comp. physiol. Psychol.*, 1958, 51, 65-67.
- LIPSITT, L. P., CASTANEDA, A., & KEMBLE, J. D. Effects of delayed reward pretraining on discrimination learning of children. *Child Developm.*, 1959, 30, 273-278.
- LOGAN, F. A. The role of delayed reinforcement in determining reaction potential. *J. exp. Psychol.*, 1952, 43, 393-399.
- LOGAN, F. A. *Incentive*. New Haven: Yale Univ. Press, 1960.
- LOGAN, F. A., BEIER, E. M., & ELLIS, R. A. Effect of varied reinforcement on speed of locomotion. *J. exp. Psychol.*, 1955, 49, 260-266.
- LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, 52, 65-70.
- LUTZ, R. E., & PERKINS, C. A time variable in the acquisition of observing responses. *J. comp. physiol. Psychol.*, 1960, 53, 180-182.
- MCGUIGAN, E. J. The effect of precision, delay, and schedule of knowledge of results on performance. *J. exp. Psychol.*, 1959, 58, 79-84.
- MAHRER, A. The role of expectancy in delayed reinforcement. *J. exp. Psychol.*, 1956, 52, 101-105.
- MATHERS, B. L. A gradient of reinforcement obtained from a complex discrimination problem. *J. comp. physiol. Psychol.*, 1953, 46, 403-406.
- MELIKIAN, L. Preference for delayed reinforcement: An experimental study among Palestinian Arab refugee children. *J. soc. Psychol.*, 1959, 50, 81-86.
- MISCHEL, W. Preference for delayed reinforcement: An experimental study of a cultural observation. *J. abnorm. soc. Psychol.*, 1958, 56, 57-61.
- MISCHEL, W. Delay of gratification, need for achievement, and acquiescence in another culture. *J. abnorm. soc. Psychol.*, 1961, 62, 543-552. (a)
- MISCHEL, W. Father-absence and delay of gratification: Cross-cultural comparisons. *J. abnorm. soc. Psychol.*, 1961, 63, 116-124. (b)
- MISCHEL, W. Preference for delayed reinforcement and social responsibility. *J. abnorm. soc. Psychol.*, 1961, 62, 1-7. (c)
- MISCHEL, W., & METZNER, R. Preference for delayed reward as a function of age, intelligence, and length of delay interval. *J. abnorm. soc. Psychol.*, 1962, 64, 425-431.
- MOWRER, O. H. *Learning theory and behavior*. New York: Wiley, 1960.
- MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, 52, 61-90.
- MYERS, J. The effects of delay of reinforcement upon an operant discrimination in the pigeon. *J. exp. Psychol.*, 1958, 55, 363-368.

- NOBLE, C. E., & ALCOCK, W. T. Human delayed-reward learning with different lengths of tasks. *J. exp. Psychol.*, 1958, 56, 407-412.
- NOTTERMAN, J. M. A study of some relations among aperiodic reinforcement, discrimination training and secondary reinforcement. *J. exp. Psychol.*, 1951, 41, 161-169.
- PERIN, C. T. The effect of delayed reinforcement upon differentiation of bar responses in white rats. *J. exp. Psychol.*, 1943, 32, 95-109. (a)
- PERIN, C. T. A quantitative investigation of the delay of reinforcement gradient. *J. exp. Psychol.*, 1943, 32, 37-51. (b)
- PERKINS, C. C. The relation of secondary reward to gradients of reinforcement. *J. exp. Psychol.*, 1947, 37, 377-392.
- PETERSON, L. R. Variable delayed reinforcement. *J. comp. physiol. Psychol.*, 1956, 49, 232-234.
- PUBOLS, B. H. Delay of reinforcement, response perseveration and discrimination reversal. *J. exp. Psychol.*, 1958, 56, 32-40.
- PUBOLS, B. H. Constant versus variable delay of reinforcement. *J. comp. physiol. Psychol.*, 1962, 55, 52-56.
- RAMOND, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, 47, 248-250.
- RENNER, K. E. Delayed reward learning as a function of drive level and cue distinctiveness. Unpublished doctoral dissertation, Northwestern University, 1961.
- RENNER, K. E. Influence of deprivation and availability of goal box cues on the temporal gradient of reinforcement. *J. comp. physiol. Psychol.*, 1963, 56, 101-104.
- RIEBER, M. The effect of CS presence during delay and reward on the speed of an instrumental response. *J. exp. Psychol.*, 1961, 61, 290-294.
- RIESEN, A. H. Delayed reward in discrimination learning by chimpanzees. *Comp. Psychol. Monogr.*, 1940, 15, 1-54.
- ROBERTS, W. H. The effect of delayed feeding on white rats in a problem cage. *J. genet. Psychol.*, 1930, 37, 35-58.
- ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
- RYAN, F. J., & BILODEAU, E. A. Countertraining of a simple skill with immediate and 1-week delays of information feedback. *J. exp. Psychol.*, 1962, 63, 19-22.
- SALTZMAN, I. J. Delay of reward and human verbal learning. *J. exp. Psychol.*, 1951, 41, 437-439.
- SALTZMAN, I. J., KANFER, F. H., & GREEN-SPOON, J. Delay of reward and human motor learning. *Psychol. Rep.*, 1955, 1, 139-142.
- SAMS, C. F., & TOLMAN, E. C. Time discrimination in white rats. *J. comp. Psychol.*, 1925, 5, 255-263.
- SCOTT, E. D., & WIKE, E. L. The effect of partially delayed reinforcement and trial-distribution on the extinction of an instrumental response. *Amer. J. Psychol.*, 1956, 69, 264-268.
- SHILLING, M. An experimental investigation of the effect of a decrease in the delay of reinforcement upon instrumental response performance. Unpublished master's thesis, University of Iowa, 1951.
- SIDMAN, M. Delayed punishment effects mediated by competing behaviors. *J. comp. physiol. Psychol.*, 1954, 47, 145-147.
- SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, 54, 1-8.
- SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univ. Press, 1956.
- WARDEN, C. J., & DIAMOND, S. A preliminary study of the effects of delayed punishment on learning in the white rat. *J. genet. Psychol.*, 1931, 39, 455-462.
- WARDEN, C. J., & HAAS, EDNA L. The effect of short intervals of delay in feeding upon speed of maze learning. *J. comp. Psychol.*, 1927, 7, 107-115.
- WATSON, J. B. The effect of delayed feeding upon reaction. *Psychobiology*, 1917, 1, 51-60.
- WIKE, E. L., & KINTSCH, W. Delayed reinforcement and runway performance. *Psychol. Rec.*, 1959, 9, 19-28.
- WIKE, E. L., & McNAMARA, H. J. The effects of percentage of partially delayed reinforcement on the acquisition and extinction of an instrumental response. *J. comp. physiol. Psychol.*, 1957, 50, 348-351. (a)
- WIKE, E. L., & McNAMARA, H. J. Some training conditions affecting secondary reinforcement. *J. comp. physiol. Psychol.*, 1957, 50, 345-347. (b)
- WIKE, E. L., & REMPLE, R. Delayed reinforcement, selective learning and habit reversal. *Psychol. Rec.*, 1959, 9, 179-187.
- WIST, E. R. Amount, delay, and position of delay of reinforcement as parameters of runway performance. *J. exp. Psychol.*, 1962, 63, 160-166.
- WOLFE, J. B. The effect of delayed reward upon learning in the white rat. *J. comp. Psychol.*, 1934, 17, 1-21.
- WOOD, A. B. A comparison of delayed reward and delayed punishment in the formation of a brightness discrimination habit in the chick. *Arch. Psychol.*, 1933, No. 157.
- YARBROUGH, J. U. The influence of the time interval upon the rate of learning in the white rat. *Psychol. Monogr.*, 1921, 30(2, Whole No. 135).

(Received December 21, 1962)

# DISPLACEMENT ACTIVITY AND MOTIVATIONAL THEORY:

## A CASE STUDY IN THE HISTORY OF ETHOLOGY<sup>1</sup>

H. PHILIP ZEIGLER

*City College of the City University of New York*

Behaviors occurring out of their characteristic motivational context and in the midst of conflict and thwarting situations have been called "displacement activities." Such behaviors were originally explained by reference to energy models of motivation. It is now known that the nature and intensity of such activities are primarily a function of 3 sets of variables: type and intensity of peripheral stimulation, the existence of behaviors incompatible with the activity in question, and the existence and duration of states of motivational equilibrium with respect to such incompatible behaviors. The implications of these findings for motivational theory are discussed and it is suggested that drive and energy concepts no longer serve any useful function in the study of species-specific behavior.

The existence of species-specific behaviors raises three distinct sets of problems for the student of animal behavior (Lashley, 1938). The first concerns the nature of the response patterns which constitute the behaviors and the stimulus situations eliciting these response patterns. Take, for example, the instinctive behavior of nest building by birds. Construction of the nest involves the utilization by the bird of certain relatively stereotyped motor sequences, elicited by certain external stimuli. Thus nest building in the domesticated canary (Hinde, 1958) involves primarily three groups of distinct motor sequences: gathering, carrying, and building. All of these behaviors are obviously elicited by specific stimuli: gathering, by the pres-

ence of grass and/or feathers; carrying, by stimuli in the bill; building, by features of both the material and the nest site. Such sensorimotor sequences are characteristic of all species-specific behaviors and the problems they raise are primarily those of the morphology and developmental history of such stimulus-response relationships.

It was to the latter problem that the early American behaviorists addressed themselves primarily, and current research on the effects of early experience upon adult behavior (Beach & Jaynes, 1954) is a direct descendant of the types of investigations carried out by Kuo (1932), Cruze (1935), and Carmichael (1936). These studies have had as their point of reference the perennial problem of the relative contributions of innate and acquired factors to the behavioral repertoire of organisms. Accordingly, they have been aimed at elucidating the developmental history of certain of the behavior patterns which are to be found in the repertoire of adult animals. In so doing, they have contributed notably to the solution of the first of the three groups of problems raised by species-

<sup>1</sup> This paper is based, in part, upon discussions with various members of the Department of Zoology, Subdepartment of Animal Behavior, Cambridge University, during the author's tenure as a Postdoctoral Research Fellow, United States Public Health Service, 1958-60. Preparation of the manuscript was supported by Grant M-5214, National Institute of Mental Health, United States Public Health Service. For their detailed comments and criticisms of the manuscript, the author is indebted to R. A. Hinde and Hugh Rowell.

specific behavior and clarified the conceptual difficulties involved in the artificial dichotomy between the innate and the acquired (Hebb, 1953; Schneirla, 1956).

However, even when the developmental history of a given sensorimotor sequence has been sufficiently clarified, certain diverse characteristics of species-specific behavior confront us with a second group of problems. Consider once again the example of nest building. Even though the sensorimotor mechanisms involved in this behavior are presumably available throughout most of the bird's adult life, nest building takes place only at certain periods in the bird's life cycle. During these periods the bird is differentially responsive to stimuli which may have been consistently present in its environment, for example, grass and feathers, but which up to this time have been ignored. Such periodic or seasonal variations in the responsiveness of sensorimotor mechanisms are characteristic of many species-specific behaviors. Witness seasonal migrations in birds and fish, or the variations in sexual receptivity of female rats during the estrus cycle. Moreover, in addition to such relatively long-term variations, short-term variations in responsiveness are also common; that is, a constant stimulus presented at intervals to an animal does not elicit a constant response. The most obvious examples are to be found in connection with feeding and mating behavior. The animal that has eaten is no longer responsive to food, and sexual responsiveness in the male is considerably reduced following copulation. Similarly, feathers and grass may be available continuously, but nest building occurs in discrete bouts which alternate with other activities. The occurrence of nest building appears to reduce temporarily the responsiveness of

the sensorimotor mechanisms involved in the behavior.

The third set of problems raised by species-specific behavior is best illustrated by the random exploratory behavior of a hungry animal, or the activity of an animal engaged in nest-site selection. Such animals appear to be reacting to the deprivation of some stimuli, and for this reason Lashley characterized such behaviors as reactions to a deficit. Thus when an animal builds a nest it is clearly reacting to specific stimuli, but nest-site selection and the fact that the nest is built to a definite form suggests, in addition, reactions to a deficit. Similarly, at a certain stage of nest building, the canary stops gathering grass and begins to collect feathers which are used for the soft, inner lining of the nest. There are no obvious external stimuli which are adequate to account for the onset of nest building, the changeover from grass to feathers, or the termination of the nest with the achievement of a definite species-specific form.

The problems raised by variations in responsiveness and reactions to a deficit, because they relate to the arousal, direction, and termination of behavior patterns rather than to their developmental history, are often grouped under the rubric of motivational problems. Ethological theories of motivation represented an attempt to deal systematically with such problems.

#### *Drive Theory, Energy Models, and Motivation*

In endeavoring to deal with the motivational aspects of behavior—learned or species-specific—many theorists have found it necessary to postulate the existence of either single- or multiple-drive processes. In either case, the drive is viewed as activating specific behavior patterns: different degrees of activation resulting in increases or decreases in the

probability of occurrence of their correlated behavior patterns.

Implicit in many of these theories is the concept of motivational energy, which is often treated as analogous to and comparable with physical, chemical, or electrical energy. While most contemporary theories of motivation have been variants of drive theory, for certain theorists, for example, Hull, Hebb, Spence, etc., the energy analogy is merely implicit and is represented by a general-drive variable whose function it is to energize behavior. For other theorists, for example, Freud, McDougall, Lorenz, and Tinbergen the centrality of the energy concept is indicated by the use of such terms as libido, psychophysical energy, reaction-specific energy, motivational impulses, etc. These latter theorists, moreover, have made the energy analogy still more explicit by embodying it in electrical, mechanical, or hydraulic models. In these energy models of motivation (Hinde, 1960), variations in the quantity, quality, or distribution of such energy are used to account for those observed variations in the behavior of animals customarily grouped under the rubric of motivational changes.

Lorenz's theory, for example, was an attempt to account for the short-term variations in responsiveness typical of species-specific behavior. His observations that sensorimotor sequences become progressively more difficult to elicit with repeated presentations of the sign stimulus led him to suggest that these variations in responsiveness reflected variations in the level of accumulated action-specific energy in some hypothetical reservoir within the animal (Lorenz, 1950).

While such concepts might account for the motivational aspects of rather limited behavior patterns (consumma-

tory behaviors), they were patently inadequate to deal with either long-term or with periodic variations in responsiveness or with the problems raised by appetitive behaviors—those behaviors which Lashley had characterized as reactions to a deficit. It was to deal with such behaviors that Tinbergen (1951) developed his hierarchichal model of motivation. This model retained Lorenz's concept of action-specific energy but extended the concept considerably. Instead of postulating motivational energy whose action specificity was restricted to a particular, limited response pattern, he viewed motivational energy as being general to all the activities of a major instinct. He thus replaced action-specific energy with what might be termed drive-specific energy. Thus the same type of motivational energy could activate all the diverse sensorimotor mechanisms grouped by Tinbergen under the rubric of the reproductive instinct—mating, nest building, aggression, maternal behavior, etc. Each of these subinstincts, however, was also activated by motivational energy specific to it, which was accumulated in its own specific reservoir or center.

The difficulties inherent in the use of such energy models have been perceptively discussed by Lehrman (1953), Kennedy (1954), and Hinde (1959, 1960). Although none of these authors considers displacement activity in detail, it is clear from their comments that the concept of displacement activity has been among the problematic aspects of ethological theories of motivation. It is equally clear from Hinde's (1960) recent paper that the vicissitudes which this concept has undergone indicated the necessity for a revision of such theories. It is for this reason that the present paper is subtitled "A Case Study in the History of Ethology."

### *Displacement Activity*

One of the most obvious characteristics of animal behavior is its directedness. Any individual pattern of behavior normally occurs as part of a sequence of functionally related behavior patterns. It appears to serve the same ends as do the behavior patterns which precede and follow it, and to share with them a common history of environmental stimulation. The grooming of birds, for example, normally occurs after bathing, when it serves to clean and resettle the feathers. It is thus associated with a recent history of certain types of environmental stimulation, for example, water or dirt on the feathers, and we should expect it to occur only in its proper place in the functional sequence and only in the presence of such stimulation. Should grooming occur in the absence of such stimulation and in the midst of a sequence of behaviors (e.g., fighting, mating, etc.) which does not include bathing, we should be justified in considering such grooming behavior anomalous, since it appears out of its usual behavioral context.

It is thus a matter of some surprise and considerable interest to find that out-of-context activities are observed quite frequently in the behavior of animals. Thus, for example, postures and movements normally associated with feeding, grooming, nest building, or sleeping in birds are often observed to occur out of context during aggressive or sexual encounters with other birds (Andrew, 1956a; Baggerman, Baerends, Helkins, & Mook, 1956; Hinde, 1952; Lorenz, 1941; Makkink, 1931, 1936; Moynihan, 1953; Tinbergen, 1940). Similar observations have been recorded for fish (Tinbergen & van Iersel, 1947) and invertebrates (Bastock & Manning, 1955; Crane, 1957; Gordon, 1955). At the mammalian level, out-of-context behavior has been observed in rats (Bolles,

1960), cats (Armstrong, 1950), dogs (Schmidt, 1956), and wolves (Schenkel, 1947). Finally, both Lorenz (1950) and Tinbergen (1951) have suggested that such characteristic human behaviors as head scratching and tie straightening in emotional situations may be viewed as out-of-context activities (for an earlier review, see Armstrong, 1950).

The possible theoretical significance of out-of-context activities was first noted, independently, by Kortlandt (1940) and Tinbergen (1940) who pointed out certain peculiarities in the behavior patterns themselves or in the conditions under which they occurred. First, such behaviors appeared to be most frequent in situations of either conflict or thwarting. A conflict situation may be characterized by the simultaneous activation of two or more incompatible behavior patterns. Such situations occur frequently during sexual (attack, escape, copulate) or aggressive encounters (attack, escape). The various behavior patterns elicited by the situation are incompatible and the animal generally alternates among them. Thwarting situations arise when a specific response pattern is activated but the animal is prevented from performing, completing, or continuing the response. A male stickleback, for example, cannot perform coition until the female follows him to the nest, and in the presence of incompletely aroused females will often display out-of-context nest-ventilation movements.

The out-of-context behaviors observed in conflict and thwarting situations were described as being different in kind and degree from the same behaviors in their normal context. To describe these differences, Tinbergen (1940) used such terms as irrelevant, frantic, and incomplete. The term irrelevant implied that the stimuli which normally elicit the behavior in an appropriate context were

absent. Thus out-of-context feeding responses can take place in the absence of food. The terms incomplete and frantic referred to the fact that out-of-context activities are often broken off in the middle, seem hurried, and are eccentric, or imperfectly oriented. For example, bill wiping, a form of grooming in the zebra finch, is normally performed with the bird's body and head parallel to the branch or twig on which it is perched. It thus makes contact with the branch when the head is lowered and the bill wiped. By contrast, out-of-context bill wiping often takes place with the bird's body at right angles to the branch so that the bill wipes in midair (Morris, 1954).

Finally, it was noted that out-of-context activities occur most frequently when the two antagonistic response tendencies are in equilibrium, that is, when they are of (presumably) equal strength. For example, out-of-context nest digging in the stickleback was found to occur most frequently during aggressive encounters at the boundaries between two adjacent territories—a place where attack and escape behaviors would be elicited in equal degree. The incidence of such behavior could be greatly increased by crowding a large number of males into a small tank, thus increasing the number of immediately adjacent territories (Tinbergen & van Iersel, 1947).

The situations described above all share a common characteristic. In every case, a specific response tendency has been activated but its overt manifestation has somehow been blocked; either by the presence of an antagonistic response tendency, or by the absence of an appropriate stimulus situation.

On the basis of such observations, Tinbergen (1940) and Kortlandt (1940) postulated a motivational mechanism derived from and compatible with the type of drive theory current in ethology at

the time. They suggested that, in the conflict situation, the motor patterns belonging to one of the drives (e.g., escape) are incompatible with the motor patterns belonging to the other (e.g., attack) drive. In the thwarting situation, they suggested, a given drive cannot express itself fully because the required releasing stimulus is not present. In either case, the result—in terms of an energy model—is a build-up of motivational energy which cannot discharge itself through the appropriate motor channels because these channels are blocked or otherwise unavailable. Tinbergen (1951) therefore suggested that such energy finds an outlet by discharging through the center of another instinct (p. 117). Since the center through which it discharges may have no functional relation to the center which is blocked, the resultant behavior will necessarily be out of context and occur in the midst of behavior sequences to which it is irrelevant. Out-of-context behaviors were thus viewed as resulting from a sparking over (*Übersprung*) of motivational energies from the motor system of one instinct to that of another, and were termed by Tinbergen *Übersprungbewegungen* (Tinbergen, 1940). The English equivalent, displacement activities, was suggested by Armstrong (1947). Grooming, which is observed to occur in the midst of an aggressive encounter would be called displacement grooming, since it has not been elicited by the normal stimuli for grooming nor activated by a grooming drive. Rather it is viewed as being motivated by energy from fleeing and/or fighting centers which has sparked over into the grooming center. As Kortlandt (1940) put it, the displacement activity is an expression not of its own drive (autochthonous motivation) but of a strange or different drive (allochthonous motivation).

The phenomenon of displacement ac-

tivity had certain immediate implications for ethological drive theory. In Lorenz's original formulation of the theory, each fixed action pattern (sensorimotor sequence) in the animal's repertoire was presumably activated from a reservoir of motivational energy specific to that pattern. The phenomenon of displacement activity suggested that surplus energy, overflowing from a given reservoir, could activate behavior patterns not normally fed from that reservoir. Similarly, Tinbergen had originally viewed motivational impulses as being drive-specific; since they could activate only one hierarchical system of centers (e.g., that of the reproductive instinct). The discovery that motivational energy could spark over from one such hierarchical system to another, for example, from reproduction to body care, implied the existence of general as well as specific motivational energy. The theory which finally emerged (Tinbergen, 1951) combined both possibilities and provided for both general drive and multiple- or specific-drive processes. In this respect it was similar both to Freudian theory (Colby, 1955) and to much contemporary psychological theorizing.

Moreover, both Tinbergen (1951) and Lorenz (1957) suggested that the primary function of displacement activity is a cathartic one. It was viewed as permitting the resolution of conflicts between two antagonistic drives by acting as an outlet through which surplus motivational energy could blow off. Several writers (Armstrong, 1950; Bastock, Morris, & Moynihan, 1953) pointed to the obvious similarity between displacement activity in animals and a variety of neurotic and emotional behaviors in man, and in animals suffering from experimental neurosis—regression, object displacement, neurotic passivity, etc. Such similarities were of particular interest for psychoanalytic instinct theories

(Barnett, 1955; Rapaport, 1960). Furthermore, both Lorenz (1957) and Tinbergen (1951) pointed to the presence of displacement activities in man as "suggesting an instinctive organization in man basically similar to that found in animals [Tinbergen, 1951, p. 210]."

Finally, there occurred one of these processes of circular reasoning only too common in the early stages of theory development. On the one hand, the neatness and ease with which the ethological energy model accounted for the phenomenon of displacement activity was taken as an indication not only of the utility of the model, but of its validity. Conversely, the very plausibility of the model lent credence to a surplus-energy interpretation of displacement activity and militated against an explanation in more analytic terms. For this reason the surplus-energy interpretation presented by Tinbergen and Kortland in 1940 was to go essentially unchallenged for more than a decade.

### *The Causal Analysis of Displacement Activity*

The publication, in the early fifties, of comprehensive statements of ethological theory (Lorenz, 1950; Tinbergen, 1951) elicited a barrage of criticisms from both psychologists (Lehrman, 1953) and ethologists (Hinde, 1959; Kennedy, 1954). These criticisms were instrumental in stimulating more detailed investigations of the factors underlying displacement activity. The results of these investigations cast serious doubt upon the motivational uniqueness of displacement activity.

It became apparent, for example, that under certain circumstances displacement activities need not be incomplete or imperfectly oriented but might be indistinguishable in form or orientation from the same behaviors in context. Thus displacement grooming in the

zebra finch (Morris, 1954) might include the whole sequence of grooming movements—preening, stretching, shaking, scratching—used by the bird in normal grooming, while displacement feeding in this species could range from incomplete pecking at the floor to actual feeding in which the bird swallowed food. For additional examples see Armstrong (1950) and Andrew (1956b).

A second distinguishing characteristic of displacement activity had been its supposed irrelevance. Two assumptions appear to underlie the use of this term. First, the behavior was assumed to be independent of any other ongoing activity of the animal. However, it has been shown that the occurrence of one displacement activity rather than another in a given situation appears to be related to the animal's posture (Lorenz, 1957; Tinbergen, 1952). Secondly, it was assumed that the stimuli which normally elicit the activity were absent. This assumption too proved to have been unjustified, since a large body of evidence indicated that the nature of the displacement activity shown depended upon or was strongly influenced by the external stimulus situation. For example, displacement feeding in the turkey (Räber, 1948) might take the form of either eating or drinking depending upon the relative availability of food and water. Displacement feeding in the great tit (Hinde, 1952) might take the form either of turning over leaves, or pecking at the ground, depending upon whether the bird were perched in a tree or on the ground. Similarly, in the courtship of the zebra finch (Morris, 1954) the bird might show displacement feeding, if food were available; if a female were available (during a fight with another male) displacement mounting might take place, that is, mounting without either preliminary courtship or succeeding copulation. If neither of these alternatives were

available, the bird might engage in displacement grooming or displacement sleeping.

Finally, Morris (1956) and Andrew (1956b) pointed out that conflict situations in birds and mammals are accompanied by a wide variety of autonomic responses including respiratory, circulatory, and thermoregulatory changes. These, in turn, would provide a host of stimuli arising from the feathers, blood vessels, and skin which might be expected to elicit a variety of grooming activities.

With the gradual accumulation of such observations it was becoming apparent that displacement activity is not irrelevant but is clearly causally related to the ongoing activities of the animal and/or the nature of the external stimulus situation. Nevertheless, a careful analysis of the motor patterns utilized as displacement activities indicates that there is some sense in which these patterns are unique.

Of all the motor patterns available in the animal's repertoire, only a limited number are ever observed to occur as displacement activities. The reader may have already noted that displacement grooming, feeding, and sleeping have been referred to most frequently. It is difficult to characterize such patterns except to note that they are common or readily available ones for the species in question (Lorenz, 1957). Grooming, for example, is one of the commonest behavior patterns of birds and mammals. Indeed, Bolles (1960) has reported that male rats in their home cages spend 40% of their waking hours in grooming behaviors. Similarly, oral activity, both in and out of feeding situations, is common in birds and mammals. As for relatively less-common activities (e.g., nest building, mounting, incubation), these are found to occur as displacement activities only when their threshold for occurrence

has been temporarily lowered—either by internal (seasonal) changes or by the presence of external stimuli (e.g., a mate, a nest, or eggs). In short, the motor patterns exhibited most frequently as displacement activities in conflict or thwarting situations are precisely those which, *ceteris paribus*, are likely to occur with the greatest frequency even under normal circumstances.

This generalization has been shown to hold not only for complex behavior patterns such as feeding or grooming, but for the individual responses which constitute these complex behaviors. Grooming in birds, for example, may consist of the following set of responses: head shaking, preening, feather settling, bill wiping, and scratching. By studying the temporal organization of these responses, several investigators (Andrew, 1956a; Morris, 1954; van Iersel & Bol, 1958) have shown that there is a regular sequence of such responses, with one type of response regularly following a given response and preceding another. During normal grooming in the bunting, for example, bill wiping and feather settling are seen first, followed by preening and finally by scratching. Other things being equal, some types of grooming responses are more likely to occur during normal grooming than are others. It is precisely these low threshold grooming responses which are observed to occur most frequently as displacement activities.

The sole remaining characteristics of displacement activity which require explanation are the occurrence of the behaviors out of their normal contexts and in the midst of situations involving conflicting or thwarted response tendencies. Why, in addition, does displacement activity in conflict situations occur most frequently when the antagonistic response tendencies appear to be of equal strength? These characteristics consti-

tuted the crux of the problem of displacement activity and as the evidence began to accumulate several alternatives to the original surplus energy hypothesis began to emerge.

In a study of normal and displacement grooming in birds, Andrew (1956a) observed that not only was grooming one of the commonest activities in the birds repertoire, it was also one of the most easily delayed, suppressed, or interrupted activities. If a bathing bird were frightened, for example, it would flee in the midst of bathing without responding to its wet plumage by a burst of grooming activity. Andrew further observed that when grooming did occur, it was most likely to take place at a time when the animal was changing over from one activity to another—an observation since confirmed and extended to the rat by Bolles (1960). Thus it might occur before and after bursts of locomotion, or after a sexual display but before nest-building activity.

On the basis of his observations, Andrew (1956a) presented a hypothesis which accounted for the occurrence of grooming as a displacement activity during conflict situations. He suggested that the peripheral stimuli which elicit grooming are continually present but that grooming is easily suppressed by other activities. In conflict situations, when a response tendency such as escape, which normally takes precedence over grooming, is blocked by an equally strong competing response tendency such as attack, grooming may occur in response to the usual peripheral stimuli. There are then two types of causal factors underlying displacement grooming—direct factors and indirect factors. Direct factors would include the presence of the peripheral stimuli normally eliciting the behavior, while the indirect factor was the extent to which other

activities, incompatible with grooming, are present.

Bindra (1959) independently suggested that displacement activity might be explained by reference to three variables: arousal level, habit strength, and sensory cues. He postulated that animals in conflict or thwarting situations are characterized by increased levels of arousal, and that in this condition the activities which are most likely to occur are those "which are prepotent in the animal's repertoire." The specific activity which occurs will be determined by the nature of the peripheral stimuli present.

A third hypothesis (the disinhibition hypothesis) was suggested by van Iersel and Bol (1958) and Sevenster (1960, 1961). It assumed that excitatory factors are constantly present in some degree for all behavior patterns, but that inhibitory relationships exist among various behavior patterns such that the activation of one such pattern suppresses the occurrence of others. In conflict situations, the antagonistic response tendencies tend to inhibit each other thus decreasing the inhibition each alone exerts on other activities. The degree of disinhibition, together with the extent to which internal factors and external factors (peripheral stimuli) are present for other activities determines which of these activities will be shown as a displacement activity.

All three of these hypotheses represent a considerable advance on the original "surplus" hypothesis. They all discard the untenable distinction between autochthonous and allochthonous motivation; they recognize the role of the stimulus situation and they describe the occurrence of displacement activity in terms of the probability of occurrence of other behavior patterns. However, the Bindra (1959) hypothesis involves the postulation of an intervening varia-

ble of arousal level; a concept whose formal properties have frequently been analogized to those of a general-drive concept (Hebb, 1955; Hinde, 1959). Similarly, both versions of the disinhibition hypothesis are phrased in terms of drives and centers. Moreover, inhibition and disinhibition are simply more elaborate descriptive terms and their use in this context is tantamount to a restatement of the observation that grooming occurs more frequently in one situation than in another.

The problem of displacement activity has been considerably clarified by a recent experimental investigation of normal and displacement grooming in the chaffinch (Rowell, 1961). Rowell achieved a considerable degree of control of the relevant variables, of which the most important are the existence of a conflict situation and the presence of the peripheral stimuli eliciting grooming. Such stimulation was produced either by spraying the birds with a fine water spray, or by using a sticky birdseed which produced grooming responses restricted to the bill (bill wiping). Two kinds of approach-avoidance conflict situations were used, both of which could be varied in intensity.

By varying the amount and/or type of peripheral stimulation, Rowell was able to affect both the composition (body grooming versus bill wiping) and frequency of grooming responses in both normal and conflict situations. He concluded that the pattern and intensity of peripheral stimulation are the direct causal factors underlying grooming in all situations, thus confirming the first part of Andrew's (1956a) hypothesis.

He further demonstrated that the frequency of grooming varies inversely with the frequency of other activities such as locomotion and that grooming is possible only in the pauses between other activities. Furthermore, grooming does

not occur in all such pauses, but only in those above a certain minimum length (which is always longer than the duration of the actual motor patterns involved in grooming). Finally, the different types of grooming responses take different lengths of time; the most common being of short duration and thus least susceptible to interruption by other activities. Not only the occurrence but also the composition of grooming is influenced by the length of time available. This confirms another portion of the Andrew (1956a) hypothesis which suggested that the presence or absence of behaviors incompatible with grooming constitutes an indirect causal factor underlying grooming.

One further factor remains to be considered. Displacement activity is most frequent when the conflicting response tendencies are equally balanced, that is, when the animal is in a state of motivational equilibrium with respect to these tendencies. In this situation, grooming may be expected to occur in response to the usual peripheral stimuli. The existence of such an equilibrium state is not directly observable but was inferred by Rowell from variations in the bird's behavior. Rowell reasoned that if, in an approach-avoidance situation, the animal pauses in its locomotion and then reverses its direction a state of equilibrium might be presumed to have been reached, since the approach tendency has changed into an avoidance tendency (or vice versa). If, as Andrew (1956a) had hypothesized, the presence of an equilibrium state is a major factor permitting grooming, grooming should be most frequent in those pauses involving changes of direction. This prediction was confirmed in Rowell's study and, independently, in a study by Tugendhat (1960) of approach-avoidance conflicts in sticklebacks. It therefore appears that the equilibrium state is a permissive

factor, and that variations in grooming behavior in such a situation are due simply to variations in the intensity and type of peripheral stimulation.

We are now in a position to account for all the observed characteristics of displacement grooming and to extend our explanation to other types of displacement activity. Behaviors observed to occur as displacement activities are those whose threshold for elicitation is low at the time. The most common displacement activities (feeding, grooming) are maintenance activities with very low thresholds. Other behaviors, for example, sexual and maternal behaviors, will occur as displacement activities when their thresholds have been reduced by the presence of hormones or of external stimuli such as nests, eggs, or other birds. The occurrence of displacement activity in conflict situations is due to the fact that such situations provide for the existence of equilibrium states which are characterized by the temporary absence of competing responses and the presence of pauses whose duration is long enough for the displacement activity to occur. Since different responses require different minimal lengths of time to occur, those which require long times are most likely to be interrupted by other activities. Thus the appearance of incompleteness or of frantic behavior is undoubtedly due to the fact that the duration of the equilibrium is generally too short to permit the complete occurrence of the entire behavior pattern. Finally, in thwarting situations the stimuli which would elicit normal, higher threshold responses are absent and other behaviors prepotent to some extent at the time may occur in response to their normal stimuli.

Such a causal analysis of displacement activity is a far cry from its original conceptualization as a mechanism which provided cathartic relief in neurotic con-

licts by discharging surplus motivational energy.

### *Implications for Motivational Theory*

The original attractiveness of ethological energy models of motivation was due, in part, to the ease with which they appeared to account for certain diverse motivational phenomena. The experimental analysis of instinctive behavior has now progressed to the point where alternative explanations of these phenomena are possible. Short-term variations in responsiveness, for example, may frequently be due not to the exhaustion of reaction-specific energy but to stimulus-specific habituation processes, or various types of sensory feedback (Bol, 1959; Hinde, 1959; Lehrman, 1956a). Reactions to a deficit may frequently be accounted for in terms of interactions between species-specific sensorimotor sequences and species-typical environments (Lehrman, 1956b; Van der Kloot & Williams, 1953a, 1953b) though in certain cases the past experience of the animal is also a crucial variable. Finally, it is now possible to account for displacement activity without recourse to energy models of motivation.

Such models, whether Freud's, McDougall's, Lorenz', or Tinbergen's, have been inextricably linked with drive constructs, so that in discarding the energy aspects of the models we raise the possibility of discarding the drive constructs as well. However, it may be advantageous at the outset to distinguish between specific- and general-drive constructs and deal with them separately.

The term specific drive often reflects a confusion between functional and causal approaches to the analysis of behavior. On the one hand, such concepts are used as functional labels to categorize and describe groups of behavior patterns which appear to share a common function. The diverse activities involved

in nest building, since they eventuate in a completed nest, are lumped together under the rubric of a nest-building drive or under the broader rubric of a maternal drive. However, since the activities involved in nest building appear to fluctuate together in time it is often assumed that they must share a common causal mechanism. Hence the concept of a specific drive is often invoked causally as an explanation for the spontaneous occurrence, temporal organization, and goal directedness of the behavior. It would obviously be advantageous to discard such constructs and replace them with explanations of the type exemplified by the causal analysis of displacement activity. Fortunately, such analyses are rapidly becoming available for almost every aspect of species-specific behavior including sexual behavior in mammals (Beach, 1956), incubation behavior in birds (Baerends, 1959), maternal behavior in birds and mammals (Lehrman, 1961), nest building in birds (Hinde, unpublished<sup>2</sup>) and cocoon spinning in silkworms (Van der Kloot & Williams, 1953a, 1953b), hunger in mammals (Miller, 1957; Teitelbaum, 1962) and in the blowfly (Dethier & Bodenstein, 1958), exploratory behavior and play (Berlyne, 1960; Welker, 1961). The existence of such causal analyses has rendered the construct of specific drive superfluous.

Perhaps the most ubiquitous of the constructs invoked by motivational theorists is that of a general drive which functions solely to energize behavior. Although such a concept does not form an explicit part of either Lorenz' or Tinbergen's theory, it was implicit in the original formulation of the surplus hypothesis of displacement activity. For psychological theorists the concept of a general drive has occupied a central

<sup>2</sup> Conference on Sex and Behavior held in 1961-62 in Berkeley, California.

place in learning theory since Hull, even though the behavioral evidence customarily cited in its support has been frequently considered of dubious validity or doubtful relevance (Bolles, 1958; Estes, 1958). The recent emergence of the concept of arousal within motivational theory represents an extension of the general-drive concept to problems other than those involved in learning—most notably to those of emotional behavior.

The term arousal was originally used in conjunction with physiological studies of a so-called arousal or activating system located in the brain stem. With the publication of papers by Duffy (1957), Lindsley (1951), Hebb (1955), and Malmö (1957), the implications of these physiological findings were explored by psychologists and the concept of arousal gained widespread currency (Berlyne, 1960; Bindra, 1959). As currently employed the term is indistinguishable in its functional properties from the concept of an energizing general drive. Some theorists have even identified general drive with the activity of the so-called nonspecific activating system of the brain stem (Hebb, 1958; Lindsley, 1957).

Such terms as arousal level and nonspecific activating system share an important feature with other drive constructs in that they have all the earmarks of what may be termed a blanket variable. They are invoked to account for a variety of diverse effects whose only common characteristic may be the fact that they all fluctuate together to some extent. Moreover, a concept of motivational energy is implicit in the arousal construct so that it is marred by the same conceptual difficulties inherent in general drive constructs. These conceptual difficulties, and certain recent developments in psychology and physi-

ology suggest the need for a re-examination of the arousal concept.

The validity of the physiological evidence upon which the concept rests is currently being called into question. A recent group of studies (Adametz, 1959; Chow & Randall, 1960; Doty, Beck, & Kooi, 1959; Kreindler, Unger, & Volanskii, 1959) indicates that the integrity of the midbrain reticular formation is not an indispensable precondition for the existence of either physiological or behavioral arousal or for the occurrence of emotional behavior or learning. Studies by Sprague, Chambers, and Stellar (1961) suggest that affective responses to sensory stimuli may be mediated by structures other than the nonspecific reticular system. Detailed analyses of brain-stem organization indicate that the so-called activation effects of reticular stimulation are not the result of a unitary arousal system but are mediated by the effects of such stimulation upon a variety of diverse physiological mechanisms (Dell, 1958; Fuster, 1961). On the basis of a comprehensive review of the anatomy and physiology of the brain-stem reticular formation, Rossi and Zanchetti (1957) concluded that this structure was unlikely to act indiscriminately as a sheer energizer.

Bindra (1959), noting that the various indices of arousal do not correlate very satisfactorily, has suggested that it may be premature to assume that these indices are representative of some unitary arousal process. Despite this caveat, explanations of emotional behavior in arousal terms continued to multiply. Indeed, it was because conflict and thwarting situations are customarily viewed as emotional situations that the concept of arousal played a central role in Bindra's hypothesis of displacement activity. As we have seen, such a concept has proved superfluous for the explanation of displacement activity and

recent studies by Schachter and his associates suggest that an arousal construct may be equally superfluous for the understanding of other types of emotional behaviors as well (Schachter & Singer, 1962).

Finally, it has been suggested that given the fact that the central nervous system is continuously active it may not be theoretically necessary to postulate an energizing factor (Hinde, 1960). Several theorists have found it possible to discard general-drive constructs entirely and present alternative approaches to the motivational problems raised by learned and species-specific behaviors. The essential feature of all these approaches is the consideration that the occurrence of a specific response in a given situation is the outcome of its successful competition with other responses which themselves have a certain probability of occurrence in that situation (Bindra, 1961; Estes, 1958; Hinde, 1959). The similarity of this formulation to those proposed by Andrew and Rowell is obvious.

One final implication of the causal analysis of displacement activity is worth considering. It will be recalled that the ethological investigators concentrated their attention not only upon the specific response under investigation (e.g., displacement grooming responses) but also upon the general behavioral context within which these responses occur. Now typically, as Bindra (1961) has pointed out, psychologists restrict their attention to the defined responses under investigation (e.g., running, bar pressing, key pecking) and ignore the matrix of general activity components (e.g., sniffing, grooming, walking, sitting) within which the defined response takes place. Bindra has convincingly demonstrated that a detailed consideration of such background behavior may shed considerable light upon the conditions

underlying the emergence of the specific responses themselves. Accordingly, it is to be hoped that in the future such a microanalysis of behavior will become a more widely-accepted technique in the study of both species-specific behavior and learning.

## REFERENCES

- ADAMETZ, J. H. Rate of recovery in cats with rostral reticular lesions. *J. Neurosurg.*, 1959, 16, 85-98.
- ANDREW, R. J. Normal and irrelevant toilet behaviour in *Emberiza* spp. *Brit. J. anim. Behav.*, 1956, 4, 85-91. (a)
- ANDREW, R. J. Some remarks on behaviour in conflict situations, with special reference to *Emberiza* spp. *Brit. J. anim. Behav.*, 1956, 4, 41-45. (b)
- ARMSTRONG, E. A. *Bird display and behavior*. London: Cambridge Univer. Press, 1947.
- ARMSTRONG, E. A. The nature and function of displacement activities. *Symp. Soc. Exp. Biol.*, 1950, 4, 361-386.
- BAERENDS, G. P. The ethological analysis of incubation behavior. *Ibis*, 1959, 101, 357-368.
- BAGGERMAN, B., BAERENDS, G. P., HELKENS, H. S., & MOOK, J. M. Observations on the behavior of the black tern, *Chlidonias n. niger* (L.), in the breeding area. *Ardea*, 1956, 44, 1-71.
- BARNETT, S. A. Displacement behavior and psychosomatic disorder. *Lancet*, 1955, 2, 1203-1208.
- BASTOCK, M., & MANNING, A. The courtship of *Drosophila melanogaster*. *Behaviour*, 1955, 8, 85-111.
- BASTOCK, M., MORRIS, D., & MOYNIHAN, M. Some comments on conflict and thwarting in animals. *Behaviour*, 1953, 6, 66-84.
- BEACH, F. A. Characteristics of masculine "sex drive." In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956*. Lincoln: Univer. Nebraska Press, 1956.
- BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, 51, 239-263.
- BERLYNE, D. E. *Conflict, arousal and curiosity*. New York: McGraw-Hill, 1960.
- BINDRA, D. *Motivation: A systematic reinterpretation*. New York: Ronald Press, 1959.
- BINDRA, D. Components of general activity and the analysis of behavior. *Psychol. Rev.*, 1961, 68, 205-215.
- BOL, ANGELA. A consummatory situation. The effect of eggs on the sexual behavior of the

- male three-spined stickleback (*Gasterosteus aculeatus*). *Experientia*, 1959, 15, 115-117.
- BOLLES, R. C. The usefulness of the drive concept. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.
- BOLLES, R. C. Grooming behavior in the rat. *J. comp. physiol. Psychol.*, 1960, 53, 306-310.
- CARMICHAEL, L. A re-evaluation of the concepts of maturation and learning as applied to the early development of behavior. *Psychol. Rev.* 1936, 43, 450-470.
- CHOW, K. L., & RANDALL, W. Learning and EEG studies of cats with lesions in the reticular formation. Paper read at First Annual Meeting Psychonomic Society, Chicago, September 2-3, 1960.
- COLBY, K. M. *Energy and structure in psychoanalysis*. New York: Ronald Press, 1955.
- CRANE, JOCELYN. Basic patterns of display in fiddler crabs (*Ocypodidae*, genus *Uca*). *Zoologica, N. Y.*, 1957, 42, 69-83.
- CRUZE, W. W. Maturation and learning in chicks. *J. comp. Psychol.*, 1935, 19, 371-409.
- DELL, P. C. Some basic mechanisms of the translation of bodily needs into behavior. In, *Neurological basis of behavior*. London: Churchill, 1958.
- DETHIER, V. G., & BODENSTEIN, D. HUNGER in the blowfly. *Z. Tierpsychol.*, 1958, 15, 129-140.
- DOTY, R. W., BECK, E. C. & KOCI, K. A. Effects of brain-stem lesions on conditioned responses in cats. *Exp. Neurol.* 1959, 1, 360-385.
- DUFFY, ELIZABETH. The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, 1957, 64, 265-275.
- ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.
- FUSTER, J. M. Excitation and inhibition of neuronal firing in visual cortex by reticular stimulation. *Science*, 1961, 133, 2011-2012.
- GORDON, H. R. S. Displacement activities in fiddler crabs. *Nature*, 1955, 176, 356.
- HEBB, D. O. Heredity and environment in mammalian behavior. *Brit. J. anim. Behav.*, 1953, 1, 43-47.
- HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, 62, 243-254.
- HEBB, D. O. *A textbook of psychology*. Philadelphia, Pa.: Saunders, 1958.
- HINDE, R. A. The behavior of the great tit (*parus major*) and some other related species. *Behaviour*, 1952, Suppl. No. 2.
- HINDE, R. A. The conflict between drives in the courtship and copulation of the chaffinch. *Behaviour*, 1953, 5, 1-31.
- HINDE, R. A. The nest building behavior of domesticated canaries. *Proc. Zool. Soc. Lond.*, 1958, 131, 1-48.
- HINDE, R. A. Unitary drives. *Brit. J. anim. Behav.*, 1959, 1, 130-141.
- HINDE, R. A. Energy models of motivation. *Symp. Soc. Exp. Biol.*, 1960, 14, 199-213.
- KENNEDY, J. S. Is modern ethology objective? *Brit. J. anim. Behav.*, 1954, 2, 12-19.
- KORTLANDT, A. Wechselwirkung zwischen Instinkten. *Arch. Neerl. Zool.*, 1940, 4, 442-520.
- KREINDLER, A., UNGER, L., & VOLANSKII, D. Effects of a circumscribed lesion of reticular formation in brain stem on higher nervous activity of dogs. *Sechenov Physiol. J. USSR*, 1959, 45, 247-256.
- KUO, Z. Y. Ontogeny of embryonic behavior in Aves: IV. The influence of embryonic movements upon behavior after hatching. *J. comp. Psychol.*, 1932, 14, 109-122.
- LASHLEY, K. S. Experimental analysis of instinctive behavior. *Psychol. Rev.*, 1938, 45, 445-471.
- LEHRMAN, D. S. A critique of Konrad Lorenz's theory of instinctive behavior. *Quart. Rev. Biol.* 1953, 28, 337-363.
- LEHRMAN, D. S. Comparative physiology (behavior). *Annu. Rev. Physiol.*, 1956, 18, 527-542. (a)
- LEHRMAN, D. S. [On the organization of maternal behavior and the problem of instinct.] In, *L'instinct dans le comportement des animaux et de l'homme*. Paris: Masson et Cie, 1956. (b)
- LEHRMAN, D. S. Hormonal regulation of parental behavior in birds and infrahuman mammals. In W. C. Young (Ed.), *Sex and internal secretions*. Baltimore: Williams & Wilkins, 1961. Pp. 1268-1382.
- LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 473-516.
- LINDSLEY, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 44-105.
- LORENZ, K. Vergleichende Bewegungstudien an Anatinen. *J. Ornithol.*, 1941, 89, 19-29.
- LORENZ, K. The comparative method in studying innate behavior patterns. *Symp. Soc. Exp. Biol.*, 1950, 4, 221-268.
- LORENZ, K. The past twelve years in the comparative study of behavior. In G. H. Schiller (Ed.), *Instinctive behavior*. New York: International Univer. Press, 1957. Pp. 288-310.

- MAKKINK, G. F. Die Kopulation der Brandente (*Tadorna tadorna* [L.]). *Ardea*, 1931, 20, 18-22.
- MAKKINK, G. F. An attempt at an ethogram of the European avocet (*Recurvirostra avosetta* L.) with ethological and psychological remarks. *Ardea*, 1936, 31, 1-60.
- MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, 64, 276-287.
- MILLER, N. E. Experiments on motivation. *Science*, 1957, 126, 1271-1278.
- MORRIS, D. The reproductive behaviour of the zebra finch (*poephilia tuttata*) with special reference to pseudofemale behaviour and displacement activities. *Behaviour*, 1954, 6, 271-322.
- MORRIS, D. The feather postures of birds and the problem of the origins of social signals. *Behaviour*, 1956, 9, 75-114.
- MOYNIHAN, M. Some displacement activities of the black-headed gull. *Behaviour*, 1953, 5, 58-80.
- RÄBER, H. Analyse des Balzverhaltens eines domestizierten Truthans (*Meleagris*). *Behaviour*, 1948, 1, 237-266.
- RAPAPORT, D. Psychoanalysis as a developmental psychology. In B. Kaplan & S. Wapner (Eds.), *Perspectives in psychological theory: Essays in honor of Heinz Werner*. New York: International Univer. Press, 1960. Pp. 209-225.
- ROSSI, G. F., & ZANCHETTI, A. The brain-stem reticular formation: Anatomy and physiology. *Arch. Ital. Biol.*, 1957, 95, 199-435.
- ROWELL, C. H. F. Displacement grooming in the chaffinch. *Anim. Behav.*, 1961, 9, 38-63.
- SCHACHTER, S., & SINGER, J. Cognitive, social, and physiological determinants of emotional states. *Psychol. Rev.*, 1962, 69, 379-399.
- SCHENKEL, R. Ausdrucks-Studien an Wölfen. *Behaviour*, 1947, 1, 81-130.
- SCHMIDT, H. D. Das Verhalten von Haushunden in Konflikt-situationen. *Z. Psychol.*, 1956, 159, 162-245.
- SCHNEIRLA, T. C. Interrelationships of the "innate" and the "acquired" in instinctive behavior. In, *L'instinct dans le comportement des animaux et de l'homme*. Paris: Masson et Cie, 1956. Pp. 387-452.
- SEVENSTER, P. The mechanism of a displacement activity. *Arch. Neerl. Zool.*, 1960, 13, 576-579.
- SEVENSTER, P. A causal analysis of a displacement activity (fanning in *Gasteroteus Aculeatus* L.). *Behaviour*, 1961, Suppl. No. 9.
- SPRAGUE, J. M., CHAMBERS, W. W., & STELLAR, E. Attentive, affective and adaptive behavior in the cat. *Science*, 1961, 133, 165-173.
- TEITELBAUM, P., & EPSTEIN, A. N. The lateral hypothalamic syndrome. *Psychol. Rev.*, 1962, 69, 74-90.
- TINBERGEN, N. Die Übersprungbewegung. *Z. Tierpsychol.*, 1940, 4, 1-40.
- TINBERGEN, N. *The study of instinct*. Oxford: Oxford Univer. Press, 1951.
- TINBERGEN, N. "Derived" activities: Their causation, biological significance, origin, and emancipation during evolution. *Quart. Rev. Biol.*, 1952, 27, 1-32.
- TINBERGEN, N., & VAN IERSEL, J. J. A. Displacement reactions in the three-spined stickleback. *Behaviour*, 1947, 1, 56-63.
- TUGENDHAT, B. The disturbed feeding of the three-spined stickleback: I. Electric shock is administered in the food area. *Behaviour*, 1960, 16, 159-187.
- VAN DER KLOOT, W. G., & WILLIAMS, C. M. Cocoon construction by the cecropia silkworm: I. The role of the external environment. *Behaviour*, 1953, 5, 141-156. (a)
- VAN DER KLOOT, W. G., & WILLIAMS, C. M. Cocoon construction by the cecropia silkworm: II. The role of the internal environment. *Behaviour*, 1953, 5, 157-174. (b)
- VAN IERSEL, J. J. A., & BOL, ANGELA. Preening of two tern species: A study on displacement activities. *Behaviour*, 1958, 13, 1-88.
- WELKER, W. I. An analysis of exploratory and play behavior in animals. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience*. Homewood, Ill.: Dorsey Press, 1961. Pp. 175-226.

(Received February 28, 1963)

## SOCIAL-LEARNING THEORY, SOCIAL DESIRABILITY, AND THE MMPI

ALFRED B. HEILBRUN, JR.<sup>1</sup>

*University of Iowa*

Recent factor analytic interpretations of the MMPI have emphasized response sets as sources of variance, and the tendency to answer items in socially desirable or undesirable ways has been interpreted as the primary factor. It is contended that the primary factor is better interpreted as degree of psychopathology and both social-learning theory and empirical data are presented to support this view. In brief, socially desirable behaviors are typically those which conform to group standards, and the deviant behaviors which define psychopathology are usually socially undesirable. Accordingly, the tendency for MMPI scales to load inversely to social desirability on the primary factor is a consequent of the actual inverse relationship between deviancy and social acceptability of behavior.

Early in the development of the MMPI the test developers recognized that the item responses were susceptible to deception and considerable effort went into establishing special scales to detect (*L* and *F*) and correct for (*K*) dissimulation (Meehl & Hathaway, 1946). Since that time there have been a great many published studies investigating the question of how much performance variance on the MMPI can be attributed to level and type of psychopathology and how much to dissimulation. It is surprising that with as much accumulated evidence as there is, the issue is still a controversial one. In the opinion of the writer, some clarification may result from expanding the psychometric question to include the more basic issue of the relationship between socialization and psychopathology. The first purpose of this paper is to present a review of those studies in which the effect of deception upon MMPI performance has been investigated. The second goal is to present an interpretation of the old evidence from the standpoint of social-learning theory and to muster some rela-

tively new evidence in behalf of the interpretation.

### REVIEW OF THE LITERATURE

#### *Effect on MMPI Performance of Instructed Simulation*

When the MMPI has been administered with instructions to simulate normal or psychopathologic performance, inconsistent results have been obtained. Hunt (1948) found that college students were able to alter their profiles when told to fake good or fake bad. Wiggins (1959) and Wiggins and Rumlill (1959) report moderate correlations between MMPI scales and a scale derived by social-desirability role-playing instructions to college subjects. Cofer, Chance, and Judson (1949) report that pattern differences were obtained, relative to standard instructions, when college subjects were asked to simulate normal or abnormal performance. Greater shifts followed from the abnormal test-taking set. When psychiatric patients were given the MMPI under instructions to simulate normalcy, Grayson and Olinger (1957) found a general tendency to change their profile patterns in the appropriate direction. However, wide individual differences between patients in the ability to simu-

<sup>1</sup> The author would like to extend his gratitude to Leonard D. Goodstein for his helpful criticism of the proposals presented in this paper.

late normalcy were observed and only 11% obtained normal simulated profiles. Gough (1954) requested college subjects to respond to the MMPI like psychoneurotics, and he found that their responses failed to correspond to those of neurotics. Even when professionally trained clinicians were employed as subjects, Gough (1947) reports a failure to simulate psychoneurotic and psychotic profiles.

The general conclusion suggested by these studies is that MMPI responses can be modified under instructed simulation conditions so that the resultant test profiles appear more adjustive and, especially, more maladjustive; however, specific types of psychopathology are far less susceptible to simulation.

### *Special Dissimulation Scales*

Of the three MMPI scales originally devised as measures of dissimulation (Lie, F, and K), K has received the greatest research attention. The K scale was developed as a measure of test-taking defensiveness by contrasting the MMPI responses of psychopathic hospital patients who presented normal profiles with the responses of subjects from the original norm group. Five of the nine clinical scales showed improved discrimination between normal and psychiatric criterion groups when K was added as a suppressor variable.

Despite the empirical derivation of K, some studies have reported a failure to improve MMPI prediction by its use (Hunt, 1948; Schmidt, 1948; Tyler & Michaelis, 1953). Others have found K useful in detecting positive malingering (Cofer et al., 1949; Hanley, 1956) and in improving discrimination between normal and disturbed college groups (Heilbrun, 1963). Smith (1959) challenged the interpretation of K as a measure of defensiveness by finding a negative relationship between K and peer ratings

of defensiveness in his normal groups. He suggested that K should be regarded as an index of psychological health, not defensiveness, in a normal population. The positive correlation between K and a measure of self-acceptance (Zuckerman & Monashkin, 1957) can be interpreted as supporting Smith's contention. Further support for a differential interpretation of K as a function of the population tested was provided by Heilbrun (1961) who found decreasing correlations between K and an independent defensive measure with more satisfactory adjustment of the college samples studied. One clear implication is that the denial of psychopathology implicit in keyed responses on the K scale is less likely to represent defensiveness when the respondent is within the normal range of adjustment; studies employing grossly normal subjects might, consequently, find K to be a less-effective suppressor variable.

Several scales have been constructed to measure the tendency to endorse socially desirable responses on the MMPI. Hanley's Ex scale (1957) included items of average popularity and neutral in social desirability; zero to moderate correlations with the MMPI scales were found. Wiggins (1959) employed social desirability role-taking procedures to develop his Sd scale, low to moderate relationships to the MMPI scales being reported. The SD scale, rationally derived by Edwards (1957), has been repeatedly found to be highly related to scores on the MMPI (Crowne & Marlowe, 1960; Edwards, 1961; Edwards, Heathers, & Fordyce, 1960; Fordyce, 1956; Wiggins, 1959). In addition, ratings of social desirability have also been shown to be highly related to MMPI scale scores (Rosen, 1956; Taylor, 1959).

The studies involving measures of social desirability leave little room for

doubt that substantial negative covariance exists between social desirability and the MMPI. The greater the probability that a subject responds in a socially desirable direction on the MMPI, the more normal will his test profile appear. There remains, however, the question of interpretation of this psychometric relationship which will be discussed at some length later in this paper.

#### *Factor-Analytic Studies of the MMPI*

Factor-analytic studies reported before 1960 have been consistent in finding that from one to three (usually two) major factors have accounted for most of the MMPI scale interrelationships whether the samples studied were drawn from normal or neuropsychiatric populations (Abrams, 1950; Cook & Wherry, 1950; Cottle, 1950; Kassebaum, Couch, & Slater, 1959; Little, 1949; Stout, 1949; Tyler, 1951; Welsh, 1956; Wheeler, Little, & Lehner 1951; Winne, 1950). Interpretation of factor meanings has been more variable but even in this regard fair agreement exists. Factor 1 with high loadings on the *Sc* and *Pt* scales has been labeled "psychotic" (Cottle, 1950; Wheeler, 1951), general maladjustment (Welsh, 1956), "tendency to personality maladjustment" (Cook & Wherry, 1950), and "ego strength versus ego weakness" (Kassebaum et al., 1959). A more inclusive Factor 1, interpretable as a general maladjustment factor, was obtained by Stout (1949) and Tyler (1951). A second neurotic factor, with high loadings on the neurotic triad (*Hs*, *D*, and *Hy*), has also been commonly extracted (Cook & Wherry, 1950; Cottle, 1950; Tyler, 1951; Wheeler et al., 1951).

More recent factor-analytic investigations of the MMPI have proposed that the two primary factors should be interpreted in terms of the test-taking habits

of the respondent rather than in terms of adjustment level. Edwards and Heathers (1962) were able to show a correlation of  $-.985$  between the *SD* scale and loadings of the MMPI scales on the primary factor, and suggested that the major source of variance on the MMPI is the tendency to give socially desirable responses versus the tendency to give socially undesirable responses to true-false type personality items. Messick and Jackson (1961) have proposed that the first MMPI factor is better interpreted in terms of acquiescence, but Edwards and Diers (1962) have provided additional evidence that loadings on the first factor can be predicted from the proportion of items keyed for socially desirable responses on MMPI scales or from zero-order correlations of the scales with the *SD* scale. In the latest study by Jackson and Messick (1962), results were obtained which they felt substantiated Edwards' contention that the primary factor is interpretable in terms of social desirability.

#### *Implications of the Literature Review*

That performance on the MMPI can be intentionally manipulated, given simulation instructions, so that the respondent will appear poorer or better adjusted is reasonably well established, although the extent to which dissimulation affects performance under standard instructions is left as a matter for conjecture.

Scales especially devised to detect dissimulation share the common problem of distinguishing between socially desirable but factual responses of well-adjusted subjects and the desirable but false and defensive responses of psychopathologic subjects. Crowne and Marlowe (1960) have criticized the Edwards *SD* scale in this regard, since they contend the keyed responses on the *SD* scale confound socially desirable behavior and psycho-

pathologic behavior. If such confounding exists, then the correlations reported between dissimulation scales and MMPI scales are essentially uninterpretable. Covariance between both types of scales could be just as well attributed to the psychopathology implicit in each as to the effects of dissimulation upon endorsement of pathology.

Factor-analytic evidence is not sufficient to settle the question of the MMPI's vulnerability to dissimulation. The level of pathology interpretations of the major factor for one array of investigators are left to stand against the social-desirability interpretation of other investigators. The purpose of the next section is to consider the alternative interpretations of the primary MMPI factor which have been proposed (i.e., a level of adjustment factor versus social desirability), present some additional evidence, and hopefully provide a better rapprochement between personality measurement and social-learning theory.

#### SOCIAL-LEARNING THEORY, SOCIAL DESIRABILITY, AND THE MMPI

##### *Theory of Social Development*

Although it is not yet possible to talk about a single stimulus-response theory of social development, there are certain broad principles which most stimulus-response psychologists would accept. Certainly the most basic principle would be that social behaviors which characterize an individual are primarily learned responses. Regardless of which learning-theory model one accepts, there seems to be general agreement that what is learned is contingent in some way upon the rewards and punishments which accrue as a function of our acts. Thus, in his social development the child learns to perform certain social behaviors and to avoid others, depending in large measure upon the nature of reinforcement imposed by the parents. The

parents, in turn, share a common core of standards with the larger social group, and their reinforcement schedules as well as their own behaviors, which serve as a model for the child, reflect these common standards. These familial learning experiences and concurrent experiences with peers and other nonfamilial adults serve to provide the person with a relatively stable set of standards of social behavior. Some social behaviors are "bad" (premarital coitus, unprovoked aggression) and their occurrence or even the impulse to their occurrence elicits anxiety; other social behaviors are "good" (friendliness, dependability) and social rewards have been and continue to be associated with their occurrence. Still other social behaviors have been neither highly rewarded nor highly punished in the child's social history and are best described as neutral. Mussen, Conger, and Kagan (1963) list deviation from cultural expectations as one of the major sources of anxiety for the child. They state:

Every culture has an unwritten list of valued traits that it expects its members to possess. . . . The individual's concept of himself is, to a large extent, a function of how closely his characteristics approximate the valued traits. When the individual perceives a great discrepancy between his own skill, traits, and temperamental qualities, and those he feels he *should* possess, anxiety is generated. The intensity of the anxiety is related to the degree to which the person perceives himself as deviating from his own and the culture's ideal standards [p. 147].

Individual differences in personal standards exist, needless to say. Factors such as social-economic class, sex of the child, identification model for the child, and unique events introduce variable patterns of antecedent conditions for their development. Yet within a social-economic class, like the American middle class, there is also a striking communality in social standards among its members. There is general consensus as

to which behaviors are socially desirable and which are socially undesirable.

These broad outlines of social development are generally accorded the name socialization, and this aspect of the social development process involves learning the rules governing the behavior of some larger group. Social desirability of behavior, as perceived by the individual, is a function of the magnitude of positive or negative reinforcement imposed by the larger group as these have been filtered through familial and other learning experiences.

### *SD Test Response Set*

Cronbach (1946, 1950) defined a test response set as a systematic response to items independent of item content and suggested that such sets contribute principally to error variance. When Edwards (1959) developed his own Personal Preference Schedule, he accorded the tendency to endorse socially desirable characteristics the status of a response set based upon the high correlation (.87) between the mean social desirability attributed to social behaviors by college subjects and the proportion of such subjects endorsing these behaviors as self-characteristic in a "yes-no" test format (Edwards, 1953). In a more recent statement (Edwards & Walker, 1961), "simple psychometric considerations" were considered adequate to account for the *SD* scale-MMPI correlations rather than such "dynamic" interpretations as psychoticism versus normality. Characteristically, Edwards and his colleagues have refrained from going beyond the operational definition of social-desirability response set provided by the *SD* scale scores in considering performance variance on the MMPI. Other investigators (Jackson & Messick, 1958) have taken a middle ground, preferring to consider such attributes of test performance as acquiescence and

social-desirability set as "styles" of response which may have more enduring and generalized behavioral effects. A response style is not necessarily a source of testing error, since it may have criterion relevance. Still others (Crowne & Marlowe, 1960; Heilbrun, 1962; Heilbrun & Goodstein, 1961; Wiggins, 1959) have taken a stronger stand in asserting that social-desirability responding on tests does have predictive relevance to nontest behavior and should not be considered a source of testing error. Marlowe and Crowne and their associates have provided rather convincing evidence that high and low *SD* test performers do show differential sensitivity to social approval in a wide array of experimental tasks (Barthel & Crowne, 1962; Crowne & Liverant, 1963; Crowne & Strickland, 1961; Marlowe & Crowne, 1961; Strickland & Crowne, 1962).

A positive relationship must be anticipated between social desirability of responses and actual social behaviors if one concurs with the theory of social development outlined above. The correlation of .87 between social desirability and endorsement of social behaviors on a questionnaire could legitimately be interpreted as reflecting the strength of this positive relationship in middle-class college students. The crucial question here would seem to be whether bright college subjects can and will provide veridical self-descriptions when afforded the opportunity. If one believes affirmatively, then the high correlation is best interpreted as demonstrating the influence of the socialization process; otherwise, the relationship is best interpreted as a reflection of a test-induced defensiveness.

### *Social Desirability and Psychopathology*

Since psychopathological behaviors are almost invariably deviant behaviors and social standards serve to elicit con-

formity from members of a group, it follows that disordered behaviors are socially undesirable and normal behaviors are socially desirable. Szasz (1960) and Mowrer (1960) have previously emphasized the importance of deviation from ethical standards in defining mental illness.

Inspection of the 39 MMPI items included in the Edwards *SD* scale provides remarkable support for the proposed relationship between social desirability and psychopathology. Each item is keyed for the socially desirable alternative and without exception greater social desirability appears to be associated with endorsement of the more-adjustive behavior option. Crowne and Marlowe (1960) report that judges in their study also considered the socially undesirable options on the *SD* scale to be indicative of maladjustment. Perhaps more conclusive evidence can be extracted from two studies by Heilbrun (1960) and Goodstein and Heilbrun (1959). Fifteen personality variables were rated by 26

PhD psychologists with regard to their adjustive consequences in male and female college undergraduates in the former study, whereas the latter paper presents the personal (in oneself) desirability of these same behaviors as judged by college undergraduates.<sup>2</sup> Table 1 gives the rank order of the 15 variables both with regard to rated adjustment value and personal desirability. Social (in others) desirability ranks, using Edwards original statement values, are also provided for these behaviors. The rank-order correlations between rated adjustment value of these behaviors and personal desirability is .82; the adjustment-social desirability correlation is .78. Both rho values are highly significant ( $t = 5.20$  and  $4.48$ , respectively,  $df = 13$ ,  $p < .01$ ). These data

<sup>2</sup> The personal desirability values assigned to the personality variables represent the average values assigned to the nine statements measuring each variable. The statement values were obtained by scaling procedures identical to those used by Edwards in estimating the social desirability values of the statements.

TABLE 1  
RATED ADJUSTMENT VALUE, PERSONAL DESIRABILITY, AND SOCIAL DESIRABILITY  
OF 15 PERSONALITY VARIABLES

Personality variable	Adjustment		Personal desirability	Social desirability
	Value <sup>a</sup>	Rank	Rank	Rank
Achievement	2.75	2		
Deference	2.32	8	1	3
Order	2.54	4.5	11	9
Exhibition	1.77	11	8	5.5
Autonomy	1.70	12	12	14
Affiliation	2.88	1	9	12
Intracception	2.33	7	2	1
Succorance	1.54	13	10	8
Dominance	2.19	9	13	10
Abasement	1.25	15	6	13
Nurturance	2.49	6	14	11
Change	1.96	10	3	2
Endurance	2.54	4.5	5	4
Heterosexuality	2.58	3	4	5.5
Aggression	1.30	14	7	7
			15	15

<sup>a</sup> Values may range between 1.00 (highly maladjustive) and 3.00 (highly adjustive).

show a positive and substantial relationship between what college students consider desirable behavior and what experts consider adjustive behavior.

It has been argued (Wiener, Blumberg, Segman, & Cooper, 1959) that even expert judges in a rating task like that used in the Heilbrun study simply employ stereotypes of social acceptability of behavior in making their ratings rather than the adjustive consequences of demonstrating these classes of behaviors. If so, the correlations of .82 and .78 would be more parsimoniously considered coefficients of agreement between two ratings of desirability. It should be noted, however, that in the Heilbrun study, maladjusted college students were also psychometrically compared to adjusted students on measures of the same 15 personality variables, using the Adjective Check List (Gough & Heilbrun, in press). The personality differences between the adjusted and maladjusted groups corresponded with what would be predicted from the adjustment ratings of the experts. As an illustration, of the 17/30 male and female scale comparisons for which significant group differences were obtained, the 10 traits which were more characteristic of the adjusted groups had a mean-rated adjustment score of 2.59. The seven variables more characteristic of the maladjusted groups showed a mean score of 1.51. Since the adjustment scores could range only from 1 (maladjustive) to 3 (adjustive) and there was no overlap between the two adjustment-score distributions, the correspondence between expert judgment and empirical finding is striking. These data suggest that the adjustment ratings were in accord with actual personality correlates of adjustment level rather than responses to desirability stereotypes.

### *Social Desirability and the MMPI as a Measure of Psychopathology*

The clinical scales of the MMPI were developed by empirically determining differences in item performance between clearly pathological groups and normals. Accordingly, it has been assumed that an appreciable amount of the performance variance on the test is attributable to degree and type of psychopathology. It is appropriate at this point to reconsider the basic question of whether the major source of variance for the MMPI is attributable to psychopathology or to social-desirability response tendencies.

Heilbrun (1963), in a study of an optimal *K*-weighting system for prediction of maladjustment in a college population, determined the biserial correlations between each of the 10 MMPI clinical scales and membership in normal or psychologically disturbed groups. Heine-man (1952), using college subjects, has provided desirability values for the MMPI items from which mean desirability values have been determined for the same clinical scales. Table 2 presents the rank order of these 10 scales from least to most socially desirable and two sets

TABLE 2  
SOCIAL DESIRABILITY RANK AND DISCRIMINATORY POWER OF THE 10 MMPI CLINICAL SCALES

MMPI scale	SD rank <sup>a</sup>	Biserial $r^b$		Biserial $r^c$	
		$r$	Rank	$r$	Rank
<i>Hs</i>	2	.26	6	.38	1
<i>D</i>	4	.48	1	.31	2
<i>Hy</i>	7	.38	4.5	.29	4
<i>Pd</i>	5	.40	3	.20	6
<i>Mf</i>	10	.12	10	.17	7
<i>Pa</i>	8	.24	7.5	.12	9
<i>Pt</i>	3	.38	4.5	.25	5
<i>Sc</i>	1	.42	2	.30	3
<i>Ma</i>	9	.13	9	.04	10
<i>Si</i>	6	.24	7.5	.16	8

<sup>a</sup> Least to most.

<sup>b</sup> Psychiatric cases versus normals.

<sup>c</sup> Counseling cases versus normals.

of biserial correlations ranked in order of magnitude. One set represents the ordered average correlations for college males and females, where the discrimination was between normal subjects ( $N = 900$ ) and acutely disturbed psychopathic hospital patients ( $N = 100$ ). Averaging was performed to be consistent with the social-desirability rankings which were obtained independent of sex. The second set of correlations represents the extent to which the clinical scales discriminated male college normals ( $N = 270$ ) from maladjusted counseling service clients ( $N = 30$ ). Maladjusted females were not included because the MMPI proved to be nonpredictive for this type of subject. The rank-order correlation between social desirability and the discriminatory power of the clinical scales was .75 for the acutely disturbed versus normal subjects and .77 for the maladjusted male versus normal subjects. Both rho values were significant ( $t = 3.21$  and  $3.41$ ,  $df = 8$ ,  $p < .02, .01$ ). The more socially undesirable the keyed responses on a scale of the MMPI, the more successfully that scale distinguished the self-descriptions of psychologically disturbed and normal subjects. Since the pathological groups invariably received the higher scale scores, it follows that they were endorsing the keyed, socially undesirable, and psychopathological behaviors more frequently. Therefore, the more socially undesirable the behaviors endorsed by an individual, the more likely he will evidence a manifest psychological disturbance.

That psychologically disturbed people tend to admit to being psychologically disturbed by their responses to the MMPI is an empirically demonstrated fact. That their admitted psychologically disturbed behavior is socially undesirable behavior is also beyond question. There remains, however, one miss-

ing link before this chain of reasoning will reach back to social-learning theory. Need the behavioral self-descriptions obtained on the MMPI bear any relationship to the same behaviors as they occur or have occurred in the natural (non-test) environment of the individual? If a person answers "false" to the item, "I loved my mother," is it any more likely that he failed to love his mother than the person who answers "true"? Hathaway and McKinley (1956) avoided this question by rendering it irrelevant. Empirical derivation of the items for the *Sc* scale required only that schizophrenics answer it differently than normals; in this case, schizophrenics answered it "false" with greater frequency. There was no need to show or even be interested in whether schizophrenics "really" failed to love their mothers more frequently than normals.

The point of view to be presented here is that given a "false" response to the mother item, two alternative answers to the question of test-response-actual-response equivalence are most reasonably entertained. One, those individuals who provide a socially undesirable "false" response to the statement, "I loved my mother," are in fact more likely to have behaved in this socially undesirable way than "true" responders; further, they are more likely to be psychologically disturbed than "true" responders. Two, those individuals who provide a socially undesirable "false" response to this item are not in fact more likely to have behaved in this particular socially undesirable way than "true" responders, but they will be characterized by other socially undesirable behaviors (such as answering "false" to the mother item); they are still more likely to be psychologically disturbed. Veridicality of response is not, therefore, a necessary assumption since either alternative is consonant with the social-learning theory

presented earlier; namely, that psychopathology represents a deviation from socially desirable standards of behavior and its corollary, the greater the psychopathology, the greater the deviation.

### Conclusions

The proposition set forth in this paper and explicated by theory and evidence is that the dimensions of psychological health and social desirability are in large measure one and the same. If true, then many of the research findings bearing upon the relationships of social desirability to psychological test performance have lacked adequate interpretation. The positive contingency of social behavior adjustment value upon the degree to which these behaviors elicit social approval has been unduly neglected, paving the way for misconstruing the  $r$ 's between test-based measures of social behavior and social desirability as necessary indications of test invalidity.

It is important to point out that some of the performance variance on personality tests is quite likely attributable to specific test-taking attitudes describable as "faking good" and "faking bad." The former attitude should lead to the endorsement of socially desirable responses and the latter to those which are socially undesirable. Thus, it becomes important to distinguish between the tendency to behave in socially desirable ways, which must be mirrored in test performance to the extent veridical self-appraisal is achieved, and the tendency to respond to test items without regard to fact but only with regard to the social appearance which the person wishes to create by his test performance. Block (1962) recently has also aligned himself behind the conclusions of this paper. After pointing out the necessary relationship between psychological health and socially desirable behavior, he distinguished between a façade or superficial brand of socially

desirable behavior and socially desirable behavior which is a correlate of adjustment. The former seems to describe the "fake good" set on a test and the latter the basic product of social learning experiences.

### REFERENCES

- ABRAMS, E. N. A comparative factor analytic study of normal and neurotic veterans: A statistical investigation of the interrelationships of intellectual and emotional factors as disclosed in the Primary Mental Abilities Examination and the Minnesota Multiphasic Personality Inventory. Unpublished doctoral dissertation, New York University, 1950.
- BARTHEL, C. E., & CROWNE, D. P. The need for approval, task categorization, and perceptual defense. *J. consult. Psychol.*, 1962, 26, 547-555.
- BLOCK, J. Some differences between the concepts of social desirability and adjustment. *J. consult. Psychol.*, 1962, 26, 527-530.
- COFER, C. N., CHANCE, JUNE, & JUDSON, A. J. A study of malingering on the MMPI. *J. Psychol.*, 1949, 27, 491-499.
- COOK, E. B., & WHERRY, R. J. A factor analysis of the MMPI and aptitude test data. *J. appl. Psychol.*, 1950, 34, 260-265.
- COTTLE, W. C. A factorial study of the Multiphasic, Strong, Kuder, and Bell inventories using a population of adult males. *Psychometrika*, 1950, 15, 25-47.
- CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 6, 475-494.
- CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.
- CROWNE, D. P., & LIVERANT, S. Conformity under varying conditions of personal commitment. *J. abnorm. soc. Psychol.*, 1963, 66, 547-555.
- CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, 24, 349-354.
- CROWNE, D. P., & STRICKLAND, BONNIE. The conditioning of verbal behavior as a function of the need for social approval. *J. abnorm. soc. Psychol.*, 1961, 63, 395-401.
- EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, 37, 90-93.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden Press, 1957.

- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1959.
- EDWARDS, A. L. Social desirability or acquiescence in the MMPI? A case study with the SD scale. *J. abnorm. soc. Psychol.*, 1961, 63, 351-359.
- EDWARDS, A. L., & DIERS, C. J. Social desirability and the factorial interpretation of the MMPI. *Educ. psychol. Measmt.*, 1962, 22, 501-509.
- EDWARDS, A. L., & HEATHERS, LOUISE B. The first factor of the MMPI: Social desirability or ego strength? *J. consult. Psychol.*, 1962, 26, 99-100.
- EDWARDS, A. L., HEATHERS, LOUISE B., & FORDYCE, W. E. Correlations of new MMPI scales with Edwards SD scale. *J. clin. Psychol.*, 1960, 16, 26-29.
- EDWARDS, A. L., & WALKER, J. N. Social desirability and agreement response set. *J. abnorm. soc. Psychol.*, 1961, 62, 180-183.
- FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, 20, 171-175.
- GOODSTEIN, L. D., & HEILBRUN, A. B. The relationship between personal and social desirability values of the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1959, 23, 183.
- GOUGH, H. G. Simulated patterns on the Minnesota Multiphasic Personality Inventory. *J. abnorm. soc. Psychol.*, 1947, 42, 215-225.
- GOUGH, H. G. Some common misconceptions about neuroticism. *J. consult. Psychol.*, 1954, 18, 287-292.
- GOUGH, H. G., & HEILBRUN, A. B. *Joint manual for the Adjective Check List and the ACL Need scales*. Palo Alto, Calif.: Consulting Psychologists Press, in press.
- GRAYSON, H. M., & OLINGER, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *J. consult. Psychol.*, 1957, 21, 73-77.
- HANLEY, C. Social desirability and responses to items from three MMPI scales: D, Sc, and K. *J. appl. Psychol.*, 1956, 40, 324-328.
- HANLEY, C. Deriving a measure of test-taking defensiveness. *J. consult. Psychol.*, 1957, 21, 391-397.
- HATHAWAY, S. R., & MCKINLEY, J. C. Scales 5 (masculinity-femininity), 6 (paranoia), and 8 (schizophrenia). In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic headings on the MMPI in psychology and medicine*. Minneapolis: Univ. Minnesota Press, 1956. Pp. 104-111.
- HEILBRUN, A. B. Personality differences between adjusted and maladjusted college students. *J. appl. Psychol.*, 1960, 44, 341-346.
- HEILBRUN, A. B. The psychological significance of the MMPI K scale in a normal population. *J. consult. Psychol.*, 1961, 25, 486-491.
- HEILBRUN, A. B. Social desirability and the relative validities of achievement scales. *J. consult. Psychol.*, 1962, 26, 383-386.
- HEILBRUN, A. B. Revision of the MMPI K correction procedure for improved detection of maladjustment in a normal population. *J. consult. Psychol.*, 1963, 27, 161-165.
- HEILBRUN, A. B., & GOODSTEIN, L. D. Social desirability response set: Error or predictor variable? *J. Psychol.*, 1961, 51, 321-329.
- HEINEMAN, C. E. A forced-choice form of the Taylor anxiety scale. Unpublished doctoral dissertation, University of Iowa, 1952.
- HUNT, H. F. The effect of deliberate deception on MMPI performance. *J. consult. Psychol.*, 1948, 12, 396-402.
- JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, 55, 243-252.
- JACKSON, D. N., & MESSICK, S. Response style on the MMPI: Comparison of clinical and normal samples. *J. abnorm. soc. Psychol.*, 1962, 65, 285-299.
- KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, 23, 226-236.
- LITTLE, J. W. An analysis of the Minnesota Multiphasic Personality Inventory. Unpublished master's thesis, University of North Carolina, 1949.
- MARLOWE, D., & CROWNE, D. P. Social desirability and response to perceived situational demands. *J. consult. Psychol.*, 1961, 25, 109-115.
- MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, 30, 525-564.
- MESSICK, S., & JACKSON, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychol. Bull.*, 1961, 58, 299-304.
- MOWRER, O. H. "Sin": The lesser of two evils. *Amer. Psychologist*, 1960, 15, 301-304.
- MUSSEN, P. H., CONGER, J. J., & KAGAN, J. *Child development and personality*. New York: Harper & Row, 1963.
- ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, 52, 151-158.
- SCHMIDT, H. O. Notes on the MMPI: The K factor. *J. consult. Psychol.*, 1948, 12, 337-342.
- SMITH, E. E. Defensiveness, insight, and the K scale. *J. consult. Psychol.*, 1959, 23, 275-277.
- STOUT, MARJORY. An analysis of the structure

- of the Minnesota Multiphasic Personality Inventory. Unpublished master's thesis, Pennsylvania State College, 1949.
- STRICKLAND, BONNIE, & CROWNE, D. P. Conformity under conditions of simulated group pressure as a function of need for social approval. *J. soc. Psychol.*, 1962, 58, 171-181.
- SZASZ, T. The myth of mental illness. *Amer. Psychologist*, 1960, 15, 113-118.
- TAYLOR, J. B. Social desirability and MMPI performance: The individual case. *J. consult. Psychol.*, 1959, 23, 514-517.
- TYLER, F. T. A factorial analysis of fifteen MMPI scales. *J. consult. Psychol.*, 1951, 15, 451-456.
- TYLER, F. T., & MICHAELIS, J. U. *K*-scores applied to MMPI scales for college women. *Educ. psychol., Measmt.*, 1953, 13, 459-466.
- WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minneapolis Press, 1956.
- WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, 15, 134-141.
- WIENER, M., BLUMBERG, A., SEGMAN, SARAH, & COOPER, A. Judgment of adjustment by psychologists, psychiatric social workers, and college students, and its relationship to social desirability. *J. abnorm. soc. Psychol.*, 1959, 59, 315-321.
- WIGGINS, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *J. consult. Psychol.*, 1959, 23, 419-427.
- WIGGINS, J. S., & RUMRILL, C. Social desirability in the MMPI and Welsh's factor scales A and R. *J. consult. Psychol.*, 1959, 23, 100-106.
- WINNE, J. F. The factorial composition of normal and neurotic responses to an adaptation of the Minnesota Multiphasic Personality Inventory. Unpublished doctoral dissertation, University of Pennsylvania, 1950.
- ZUCKERMAN, M., & MONASHKIN, I. Self-acceptance and psychopathology. *J. consult. Psychol.*, 1957, 21, 145-148.

(Received March 4, 1963)

## DOWN'S SYNDROME (MONGOLISM): THE HERITABLE ASPECTS

LISSY F. JARVIK, ARTHUR FALEK, AND W. PHILIP PIERSON  
*New York State Psychiatric Institute, New York City*

Striking technological advances in the study of human chromosomes have revealed an excess of genetic material in the cells of patients with Down's syndrome. Both the trisomic and translocation forms of mongolism are discussed, as is the assessment of morbidity risks in relatives of a given patient. The origin of the chromosomal abnormalities, and the mechanisms by which they generate the clinical picture of Down's syndrome, constitute questions for further research. When this etiological enigma is resolved optimal management of patients and their relatives will depend, even more than at present, upon the active collaboration of psychologists and psychiatrists with human geneticists.

Mental deficiency has been a scourge of mankind since early recorded history (Barr, 1904). Yet it was only recently that the mentally subnormal were assigned to diagnostically useful categories, and only during the current decade has a promising start been made toward progress from nosology to etiology. One of the early workers in the area of classification was J. Langdon Down who gave a concise clinical description of the syndrome now bearing his name and pointed to the well-known superficial similarities between the disorder and representatives of the Mongolian ethnic group. His 1866 report on the "Ethnic Classification of Idiots" (Down, 1887) appeared at the same time as Mendel's 1866 classic "Experiments in Plant-Hybridization," (Mendel, 1866; trans. in Peters, 1959), formulating the basic concepts of genetics. It was this very science of genetics which was responsible for our recently acquired understanding of certain biological phenomena underlying the genesis of Down's syndrome. Like that of Mendel's report, unknown to the scientific community for 35 years, the circulation of Down's paper was initially limited. Thus, Fraser and Mitchell (1876), 10 years later, stated that they had "endeavoured to obtain the literature re-

lating to this form of idiocy, but without success [p. 169]." In the excellent description of "this form of idiocy" Fraser meticulously recounted the symptoms observed in a 40-year-old woman, appending Mitchell's notes on 62 similar cases. Mitchell suggested that these cases could be grouped into a single category which he designated as Kalmuc idiocy because of the physical resemblance of the patients to Kalmucs (members of certain Mongolian tribes).

When Fraser read the paper at a meeting of the Medico-Psychological Association in 1875, he was roundly criticized by the chairman who stated that:

it struck him that the term Kalmuc was a very objectionable term. He thought there was nothing marked about the Kalmuc, and there was no reason for nicknaming an idiot by calling him a Kalmuc [Fraser & Mitchell, 1876, p. 161].

The controversy concerning nomenclature has persisted to this day as shown by a letter to *Lancet* in which 19 signatories advocated the abandonment of the term "mongolism" because of its "misleading connotations" (Allen et al., 1961). They did not champion a particular substitute but hoped "that agreement on a specific phrase will soon crystallize if once the term 'mongolism' has been abandoned."

The century-old argument regarding terminology is clearly of minor import when compared to questions of etiology. Yet prior to the dramatic discoveries of 1959 information concerning causative factors had progressed little beyond Mitchell's astute notations of late birth order, frequency of preceding miscarriages or prolonged periods of sterility, and advanced parental (especially maternal) age (Fraser & Mitchell, 1876; Øster, 1953, 1956; Penrose, 1934; Shuttleworth, 1895, 1909).

Various postulates regarding the etiology of Down's syndrome were discussed in detail by Benda in his well-known text (1949, 1960). While these hypotheses are far too numerous for review here (Ingalls, 1947; Warkany, 1960), it may be of interest to recall Down's (1887) own comments in this connection:

Has the nurse dosed the child with opium? Has the little one met with any accident? Has the instrumental interference which maternal safety demanded been the cause of what seems to the anxious parents a vacant future? Can it be that when away from the family attendant medicine has been injudiciously prescribed? Can, in fact, the strange anomalies which the child presents be attributed to the numerous causes which maternal solicitude conjures to the imagination, in order to account for a condition, for which any cause is sought rather than hereditary taint or parental influence [p. 211]?

The general reluctance to consider heritable aspects of mongolism was abetted by the paucity of pertinent data. Since familial aggregation of mongoloid patients tends to be rare, evidence for the operation of hereditary factors was essentially limited to the significant findings in twin studies as well as the occurrence of minor stigmata of the syndrome in the close relatives of patients (Fang, 1950; Penrose, 1954a, 1954b; Turpin & Caratzali, 1933). A comprehensive evaluation of all twin data then available led Allen and

Kallmann (1955) to conclude that genetic factors were of importance in the etiology of mongolism. As of 1957 concordance rates were close to 100% for one-egg pairs, while in two-egg pairs the morbidity risk for co-twins was less than 5% (Allen & Kallmann, 1957). The conclusions reached by Kallmann and his associates (Allen & Baroff, 1955; Allen & Kallmann, 1955; Baroff, 1958), based upon the significant difference in concordance rates, could not be corroborated at the time by examination of children born to mongoloid patients. Even today (1962), a generation after the introduction of modern chemotherapy, only 13 children (4 afflicted) are known to have been born to 12 mothers diagnosed as Down's syndrome (Forssman, Lehmann, & Thysell, 1961; Forssman & Thysell, 1957; Hanhart, 1960; Lelong, Borniche, Kreisler, & Baudy, 1949; Levan & Hsu, 1960; Mullins, Estrada, & Gready, 1960; Rehn & Thomas, 1957; Sawyer, 1949; Sawyer & Shafter, 1957; Schlaug, 1957; Stiles & Goodman, 1961; Thuline & Priest, 1961).

Nevertheless, our level of understanding of Down's syndrome today differs vastly from that of only 3 years ago. The dramatic discoveries made in 1959 can best be understood in the light of knowledge then existing concerning cellular aspects of human heredity. Although it had long been agreed that the nucleus of every cell contained the basic hereditary material, it was not until 1956 that methods became adequate for cultivating human cells and studying their chromosomes.

By 1959, a sufficient number of reports had appeared in the literature (Chu & Giles, 1959; Ford & Hamerton, 1956; Ford, Jacobs, & Lajtha, 1958; Tjio & Levan, 1956; Tjio & Puck, 1958a, 1958b) to justify the tenet that the normal human complement was 23 pairs or 46 chromosomes. Among them

are two sex chromosomes (one pair) which distinguish between the sexes. In the female these two chromosomes are the same (XX) while in the male they are different (XY). In addition to the one pair of sex chromosomes there are 22 other pairs, apparently the same for both sexes, which are known as autosomes. The 22 autosome pairs together with the sex chromosomes comprise the normal diploid complement of 46 chromosomes.<sup>1</sup> The autosome pairs have been assigned numbers from 1 to 22 according to their relative size and configuration and the standard nomenclature adopted by the original workers in the field is known as the Denver system (Denver Report, 1960).

Utilization of the newer techniques of tissue culture and cell microscopy (cytogenetics) soon led to the startling observation of an extra chromosome in patients with Down's syndrome. The first description of 47 chromosomes in this disorder was contained in the report of Lejeune, Gautier, and Turpin (1959a) who had analyzed the chromosomes of three mongoloid boys. The occurrence of the extra small chromosome was soon confirmed by Jacobs, Baikie, Court Brown, and Strong (1959) who had independently observed the same phenomenon. During the succeeding years numerous further studies of patients with Down's syndrome affirmed the characteristic number of 47, the extra small chromosome being designated as number 21 (Böök, Fraccaro, & Lindsten, 1959; German & Bearn, 1960; Hanhart, Delhanty, & Penrose, 1961; Lejeune, Gautier, & Turpin, 1959b; Makino, Honda, & Mitani, 1962; Miller, Breg, Schmickel, & Tretter, 1961). The cells of such individuals then contain three, instead of the

normal two, chromosomes number 21 and persons with this abnormality (trisomy number 21) are referred to as 21-trisomics.

The origin of the extra chromosome, the third one, is, of course, an intriguing problem. The two chromosome members of each of the 23 pairs ordinarily present are contributed by the parental germ cells (egg or sperm). As has long been known, each germ cell, or gamete, contains not the diploid number of chromosomes present in all other cells but only half as many, the haploid number. Thus only one member of each of the 23 chromosome pairs is represented in the germ cells (23 instead of 46 chromosomes). During fertilization the combination of egg and sperm, with 23 chromosomes each, reconstitutes the human diploid number of 46 in generation after generation.

The process by which cells containing only one half of each chromosome pair are formed is called meiosis and involves reduction division (reduction to one half the diploid number of chromosomes). It is beyond the scope of the present discussion to examine in detail the intricate biological processes that are encompassed by the terms of spermatogenesis and oögenesis (mitotic as well as meiotic divisions, crossing over, formation of polar bodies, etc.), but it is sufficient for our purposes to examine the final result consisting of eggs and sperm with but one member of each chromosome pair. When for some reason, as yet unknown, a chromosome pair does not separate during meiosis, there arises a germ cell with 24 instead of 23 chromosomes (Miller, Mittwoch, & Penrose, 1960). The union of this type of cell with a normal germ cell results in an individual with 47 rather than 46 chromosomes.

In cases of Down's syndrome this process (nondisjunction) affects chro-

<sup>1</sup> Prior to the above noted methodological advances the human diploid number was believed to be 48.

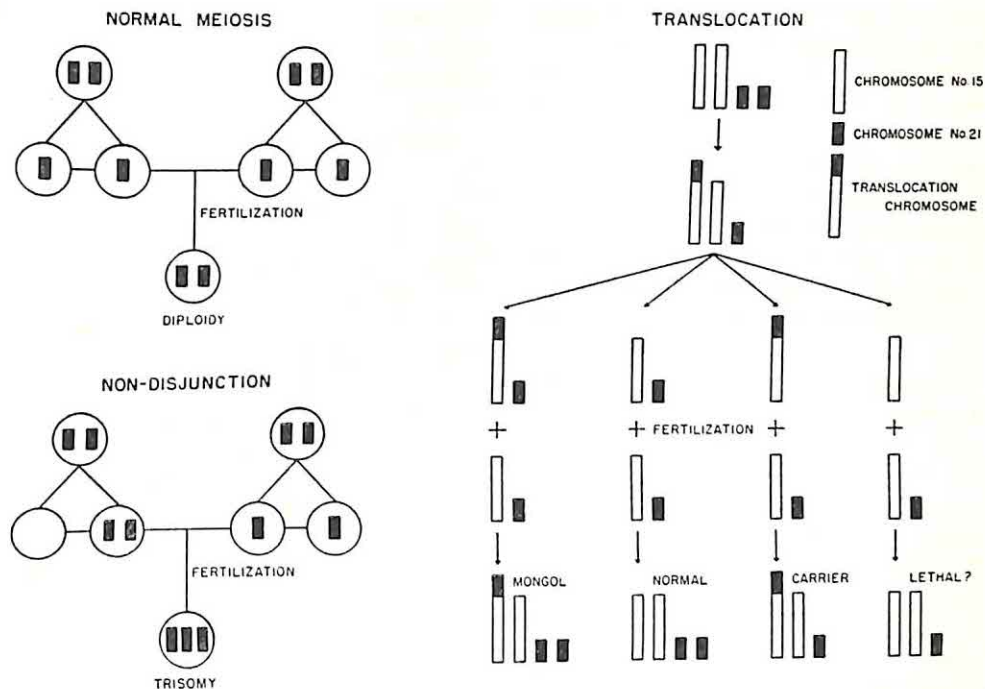


FIG. 1. Meiosis, nondisjunction and translocation. (In normal meiosis, diploid number of chromosomes is reduced to haploid number. Fertilization reconstitutes the diploid complement. In mongolism resulting from nondisjunction, chromosome pair number 21 fails to separate during meiosis and fertilization results in trisomy number 21. In mongolism due to translocation, the genetic material of chromosome number 21 is attached to another chromosome—for example, number 15. The independent assortment of chromosomes produces four possible combinations: mongol, normal, carrier, lethal.)

mosome pair number 21 (see Figure 1). Although nondisjunction is liable to occur in both egg and sperm cells, for purposes of illustration we may consider an ovum where the two members of chromosome pair number 21 remained together. When a normal sperm contributes its chromosome number 21 to such an ovum, the fertilized egg becomes trisomic for chromosome number 21. It is likely that only a small proportion of the gametes (germ cells) in a given parent are so affected and, unfortunately, we still lack the means to detect this process. It is only after a mongoloid child has been born that we know nondisjunction occurred in a parent's gametes. In the afflicted child all of the body cells are trisomic for chromosome number 21

so that the diagnosis of Down's syndrome can be made through the culture and examination of cells from a variety of sources (e.g., skin, bone marrow, or blood).

Persistent application of these cytogenetic procedures to typical patients soon revealed that some of them had 46 instead of the expected 47 chromosomes. How could the presence of Down's syndrome be reconciled with the absence of a third chromosome number 21? If the syndrome is attributed to an extra amount of genetic material from chromosome number 21, then it is reasonable to assume that in patients with 46 chromosomes this excess is attached to one of the other chromosomes. This assumption of a translocation has been

borne out by the finding of excessively long chromosomes in such patients.

The translocation most frequently reported involves the attachment of chromosome number 21 to a chromosome in the 13-14-15 group (so-called 15/21 translocation) while in other instances combinations of 22/21 and 21/21 have been observed in translocation mongolism (Atkins, O'Sullivan, & Pryles, 1962; Bieseke, Schmid, Lee, & Smith, 1962; Böök, Santesson, & Zetterqvist, 1961; Carter & Evans, 1961; Carter, Hamerton, Polani, Gunalp, & Weller, 1960; Ek, Falk, Bergman, & Reitalu, 1961; Fraccaro, Kaijser, & Lindsten, 1960; Hamerton, Cowie, Giannelli, Briggs, & Polani, 1961; Hamerton & Steinberg, 1962; Lehmann & Forssman, 1962; Macintyre, Staples, Steinberg, & Hemple, 1962; Penrose, Ellis, & Delhanty, 1960; Polani, Briggs, Ford, Clarke, & Berg, 1960; Scherz, 1962; Shaw, 1962; Turpin & Lejeune, 1961; Turpin, Lejeune, Lafourcade, & Gautier, 1959). Various families have been described in which Down's syndrome, due to translocation of chromosome number 21, has affected several members of succeeding generations. The translocation chromosome was observed not only in the affected persons but also in apparently normal parents and siblings. The difference between them reflected the total amount of chromosomal material, unaffected translocation carriers having 45, rather than 46, chromosomes. The apparently missing chromosome in these translocation carriers belonged to pair number 21 and was not really absent since it was attached to one of the other chromosomes. As illustrated diagrammatically in Figure 1, the unaffected carrier has the usual allotment of genetic material, despite the apparent shortage of one chromosome.

Although the appearance (phenotype) of translocation carriers gives no clue to

their chromosomal abnormality, present cytogenetic techniques can readily demonstrate the chromosomal anomaly. Persons who inherit the translocation chromosome from one of their parents show the abnormality in all of their body cells. This is a logical consequence of the fact that all of their cells evolved from a single fertilized ovum. Theoretically, one half of their germ cells, which carry the haploid complement, will contain 23 and the other half only 22 chromosomes. While all of the gametes with 22 chromosomes lack a morphologically discrete chromosome number 21, some of them contain the translocation. If it is true that absence of the genetic material from chromosome number 21 is incompatible with reproduction (no individual has yet been found with the equivalent of only one chromosome number 21), then there remain three kinds of germ cells which can yield viable offspring, one gamete with 22 and two with 23 chromosomes. Since the gamete with 22 chromosomes contains the translocation, fertilization by a normal germ cell will result in a phenotypically normal individual with 45 chromosomes. One of the two gametes with 23 chromosomes also contains the translocation, in addition to a separate chromosome number 21. When such a germ cell is fertilized by a normal gamete there is an excess of chromosome number 21 material which results in a child with Down's syndrome. In the third type of germ cell the chromosomes are normal in configuration as well as in number and may be expected to give rise to a normal individual. Such a person does not carry the translocation and, therefore, cannot transmit it to future generations.

Since the translocation chromosome is transmitted by unaffected carriers to two thirds of viable offspring in each succeeding generation, and since half of those will manifest Down's syndrome,

why then do we not have an abundance of case histories with affected siblings, cousins, aunts, uncles, and other relatives in collateral lines? After all, investigators have been searching for etiological factors throughout the past century and would surely not have overlooked such a family accumulation. Indeed, multiple cases have been reported for certain families occasionally, and we are now able to explain the relative rarity of such pedigrees. Obvious familial aggregation is expected only in the form of Down's syndrome due to translocation and the overwhelming proportion examined so far have shown trisomy rather than translocation.

As mentioned earlier, the trisomic form of mongolism is characterized by an extra or third chromosome number 21 which is indistinguishable from the other two chromosomes number 21 and the affected individual has a total of 47 chromosomes.

Whatever the mechanism, trisomy or translocation, there must be postulated some relationship between the excess chromosomal material and clinical symptomatology. In fact, most individuals containing cells with both the normal and deviant chromosome complements manifest relatively mild stigmata (Clarke, Edwards, & Smallpiece, 1961; Fitzgerald & Lycette, 1961; Hayashi, Hsu, & Chao, 1962; Ilbery, Lee, & Winn, 1961; Lindsten, Alvin, Gustavson, & Fraccaro, 1962; Nichols, Coriell, Fabrizio, Bishop, & Boggs, 1962; Richards & Stewart, 1962). Equating chromosomal matter with genetic information, we may say that Down's syndrome results from an extra amount of DNA (deoxyribonucleic acid), the macromolecule embodying the genetic code.

It is beyond the scope of the present discussion to enter into the physiochemical intricacies of the DNA molecule whose double helix structure was eluci-

dated by the 1962 Nobel laureates, James D. Watson, Francis H. C. Crick, and Maurice H. F. Wilkins. The Watson-Crick model (Watson & Crick, 1953) has been compared to a ladder with the legs twisted about each other. The legs of the ladder stand for the sugar-phosphate complexes and the rungs represent the paired bases (adenine-thymine and guanine-cytosine). The order in which the base pairs are arranged apparently determines the genetic code and it is easy to imagine how a surplus of coded information (excess DNA) could disturb the complex copying processes, thus interfering with normal development.

Although the excess genetic material from chromosome number 21 is now known to culminate in Down's syndrome, the conditions generating the chromosomal aberration remain largely unknown. As has been described, translocation chromosomes are transmitted according to well-established laws of heredity but the forces creating the first translocation in any family await definition. Similarly, the circumstances conducive to nondisjunction are yet to be explained. Recalling moreover that modern methods for the study of human chromosomes have been available for just over 5 years and that even these are still somewhat primitive, it may be expected that increased technical sophistication will lead to the detection of still further abnormalities in Down's syndrome. It cannot be emphasized too strongly that the discovery of chromosomal aberrations is but one step toward the detection of causal factors in mongolism. Many of the hypotheses proposed during the past century are still tenable provided we adjust our thinking to include the origin of chromosomal aberrations. In a way we are reformulating the etiological problem at a level more amenable to experimental manipulation than that of clinical symptomatology.

Among the factors held of basic importance in the past have been ionizing radiations, viruses, and senescence. Radiation-produced translocations are well known in plants and virally induced changes have been studied extensively in microbial genetics. Though the effects of senescence in man are still subject to speculation (Jarvik, 1963), observations relating an increased frequency of mongolism to advanced parental age date back to the early descriptions of the disorder as mentioned in our introductory section. On the basis of cytogenetic data it has been suggested that the relationship between parental age and Down's syndrome applies primarily to the trisomic form of mongolism resulting from nondisjunction (Carter & Evans, 1961; Hamerton, Briggs, Giannelli, & Carter, 1961; Penrose, 1961, 1962).

As a matter of fact, the association with advanced parental age has been noted for nondisjunction of chromosomes other than number 21. In Klinefelter's syndrome, for example, it is two X chromosomes which remain joined, and together with a Y chromosome give rise to abnormal males (XXY instead of XY) characterized by gynecomastia, hypogonadism, and sterility. The disorder affects about 1 in every 500 newborn boys (Bergemann, 1961; Maclean, Harnden, & Court Brown, 1961; Moore, 1959) or with approximately the frequency (1 in 650) found in Down's syndrome (Bearn & German, 1961; Penrose, 1961).

A predisposition to nondisjunction, possibly enhanced by advancing age, has been postulated with respect to Klinefelter's as well as Down's syndrome (Lubs, 1961; McKusick, 1962). Moreover, the observed number of individuals with coexistence of both aberrations, XXY and trisomy number 21 (with a total of 48 chromosomes), has been greater than that expected by chance

(Ford, Jones, Miller, Mittwoch, Penrose, Ridler, & Shapiro, 1959; Harnden, Miller, & Penrose, 1960; Lanman, Sklarin, Cooper, & Hirschhorn, 1960; Lehmann & Forssman, 1960). Chance expectation has also been exceeded in the association of mongolism with leukemia and the significance of this relationship is being explored (Buckton, Harnden, Baikie, Woods, 1961; Davis & Korst, 1960; Krivit & Good, 1957; Tough, Court Brown, Baikie, Buckton, Harnden, Jacobs, King, & McBride, 1961). Though considerably more data will be required before any definite conclusions can be reached with regard to etiology, the information gathered to date (1962) does serve to highlight numerous approaches toward penetrating the cryptic processes which lead to mongolism. With respect to parental age effects, for example, it should be profitable to examine whether nondisjunction ensues from changes intrinsic in or extrinsic to the ovum and whether abnormal ova are actually responsible for the majority of mongoloid individuals as claimed on statistical grounds. While preliminary clues may be derived from animal experiments, final answers will have to come from human material. The widespread use of oral contraceptives, resulting in the retention of large numbers of ova, otherwise discharged at monthly intervals, will provide information hitherto unavailable for human females.

In addition to defining conditions favorable or unfavorable for the development of chromosomal aberrations, there is a need for delineating the actions of the redundant genetic material. One may speculate that the excess DNA leads to an overabundance of substances ordinarily produced by the organism (normal metabolites) and that such profusion is harmful. Familiar examples of noxious effects correlated with quantitative increases in metabolites are

diabetes and thyrotoxicosis. Demonstrating this type of mechanism in Down's syndrome would encourage a causally directed, therapeutic orientation. So would the equally plausible suggestion that the anomalies seen in Down's syndrome reflect a deficiency rather than an excess of some normally required metabolites. Furthermore, the similarities between mongolism and cretinism, so striking that they create diagnostic difficulties, may also be cited in support of this view. That a deficiency hypothesis is compatible with replacement therapy is illustrated by cretinism where the timely administration of thyroid hormone may not only halt but even reverse the development of the physical and mental stigmata of the disorder.

Conceivably more than one substance and more than one mechanism may ultimately be implicated in Down's syndrome (Gershoff, Hegsted, & Trulsson, 1958; Jérôme, Lejeune, & Turpin, 1960; King, Gillis, & Baikie, 1962; O'Brien & Groshek, 1962). It would become mandatory then to distinguish for therapeutic purposes between different forms of the disorder, just as it is necessary now to differentiate for counseling purposes trisomy from translocation.

It is in the area of family counseling that a working knowledge of modern genetics may be said to constitute a sine qua none for the psychologist. Today, relatives of a mongoloid patient can be given a reasonable estimate of the probability that they themselves will have an afflicted child (ranging from less than .002% to over 30%), provided the informed counselor can obtain accurate cytogenetic data.

Because of the complex evaluation required for the correct assessment of the facts in a given counseling situation, it is imperative that anyone dispensing genetic advice be fully familiar with the current genetic literature and qualified

to interpret cytogenetic, biochemical, and routine clinicopathological reports (Hirsch & Erlenmeyer-Kimling, in press). It is hardly necessary to specify for psychologists the need for competent evaluation of psychological factors. Quotation of accurate facts and figures without consideration of their impact upon the individual seeking advice may have devastating consequences.

Since sophistication in both clinical psychology (or psychiatry) and human genetics is rarely found in a single individual, several persons, each having a portion of the requisite training, may profitably work together on a counseling problem (Kallmann, 1961; Kallmann & Rainer, 1963). With forthcoming knowledge of the etiology and pathogenesis of Down's syndrome successful therapeutic reforms will no doubt be instigated so that the heritability of this mental dysfunction may eventually be identified with curability.

#### REFERENCES

- ALLEN, G., et al. Mongolism. *Lancet*, 1961, 1, 775.
- ALLEN, G., & BAROFF, G. S. Mongoloid twins and their siblings. *Acta genet. Statist. med.*, 1955, 5, 294-326.
- ALLEN, G., & KALLMANN, F. J. Frequency and types of mental retardation in twins. *Amer. J. hum. Genet.*, 1955, 7, 15-20.
- ALLEN, G. & KALLMANN, F. J. Mongolism in twin sibships. *Acta genet.*, 1957, 7, 385-393.
- ATKINS, L., O'SULLIVAN, M. A., & PRYLES, C. V. Mongolism in three siblings with 46 chromosomes. *New Engl. J. Med.*, 1962, 266, 631-635.
- BAROFF, G. S. Current theories on the etiology of mongolism. *Eugen. Quart.*, 1958, 5, 212-215.
- BARR, M. W. *Mental defectives*. Philadelphia, Pa.: P. Blakiston's Sons, 1904.
- BEARN, A. G., & GERMAN, J. L. Chromosomes and disease. *Scient. American*, 1961, 205, 66-76.
- BENDA, C. E. *Mongolism and cretinism*. (Orig. publ. 1946) New York: Grune & Stratton, 1949.
- BENDA, C. E. *The child with mongolism*. New York: Grune & Stratton, 1960.
- BERGEMANN, E. Geschlechtschromatinbestim-

- mung am Neugeborenen. [Nuclear sex of the newborn.] *Schweiz. med. Wschr.*, 1961, 91, 292-294.
- BIESELE, J. J., SCHMID, W., LEE, C. H., & SMITH, P. M. Translocation between acrocentric chromosomes in a 46-chromosome mongoloid and his 45-chromosome mother. *Amer. J. hum. Genet.*, 1962, 14, 125-134.
- BÖÖK, J. A., FRACCARO, M., & LINDSTEN, J. Cytogenetical observations in mongolism. *Acta paediat.*, 1959, 48, 453-468.
- BÖÖK, J. A., SANTESSON, B., & ZETTERQVIST, P. Translocation heterozygosity in man. *Lancet*, 1961, 1, 167.
- BUCKTON, K. E., HARNDEN, D. G., BAIKIE, A. G., & WOODS, G. E. Mongolism and leukemia in the same sibship. *Lancet*, 1961, 1, 171-172.
- CARTER, C. O., & EVANS, K. A. Risk of parents who have had one child with Down's syndrome (mongolism) having another child similarly affected. *Lancet*, 1961, 2, 785-788.
- CARTER, C. O., HAMERTON, J. L., POLANI, P. E., GUNALP, A., & WELLER, S. D. V. Chromosome translocation as a cause of familial mongolism. *Lancet*, 1960, 2, 678-680.
- CHU, E. H. Y., & GILES, N. H. Human chromosome complements in normal somatic cells in culture. *Amer. J. hum. Genet.*, 1959, 11, 63-79.
- CLARKE, C. M., EDWARDS, J. H., & SMALLPIECE, E. V. 21 trisomy/normal mosaicism in an intelligent child with mongoloid characteristics. *Lancet*, 1961, 1, 1028-1030.
- DAVIS, J. H., & KORST, D. R. Mongolism and leukemia. *Med. Bull. U. Mich.*, 1960, 26, 6-11.
- DENVER REPORT. A proposed standard system of nomenclature of human mitotic chromosomes. *Amer. J. hum. Genet.*, 1960, 12, 384-388.
- DOWN, L. J. Observations on an ethnic classification of idiots. In, *Mental afflictions of childhood and youth*. (Orig. publ. 1866) London, England: Church, 1887. Pp. 210-217.
- EK, J. I., FALK, V., BERGMAN, S., & REITALU, J. A male mongoloid with 46 chromosomes. *Lancet*, 1961, 2, 526-527.
- FANG, T. C. The third interdigital patterns on the palms of the general British population, mongoloid and non-mongoloid mental defectives. *J. ment. Sci.*, 1950, 96, 780-787.
- FITZGERALD, P. H., & LYCETTE, R. R. Mosaicism in man, involving the autosome associated with mongolism. *Heredity*, 1961, 16, 509-512.
- FORD, C. E., & HAMERTON, J. L. The chromosomes of man. *Nature, London*, 1956, 178, 1020-1023.
- FORD, C. E., JACOBS, P. A., & LAJTHA, L. G. Human somatic chromosomes. *Nature, London*, 1958, 181, 1565-1568.
- FORD, C. E., JONES, K. W., MILLER, O. J., MITTWOCH, U., PENROSE, L. S., RIDLER, M., & SHAPIRO, A. The chromosomes in a patient showing both mongolism and the Klinefelter syndrome. *Lancet*, 1959, 1, 709-710.
- FORSSMAN, H., LEHMANN, O., & THYSELL, J. Reproduction in mongolism: Chromosome studies on re-examination of a child. *Amer. J. ment. Defic.*, 1961, 65, 495-498.
- FORSSMAN, H., & THYSELL, T. A woman with mongolism and her child. *Amer. J. ment. Defic.*, 1957, 62, 500-503.
- FRACCARO, M., KAIJSER, K., & LINDSTEN, J. Chromosomal abnormalities in father and mongol child. *Lancet*, 1960, 1, 724-727.
- FRASER, J., & MITCHELL, A. Kalmuc idiocy: Report of a case with autopsy, with notes on sixty-two cases. *J. ment. Sci.*, 1876, 22, 161-162, 169-179.
- GERMAN, J. L., & BEARN, A. G. Family studies in the chromosomal complement of mongols. *J. clin. Investig.*, 1960, 39, 989-990.
- GERSHOFF, S. N., HEGSTED, D. M., & TRULSON, M. F. Metabolic studies of mongoloids. *Amer. J. clin. Nutr.*, 1958, 6, 526-530.
- HAMERTON, J. L., BRIGGS, S. M., GIANELLI, F., & CARTER, C. O. Chromosome studies in detection of parents with high risk of second child with Down's syndrome. *Lancet*, 1961, 2, 788-791.
- HAMERTON, J. L., COWIE, V. A., GIANELLI, F., BRIGGS, S. M., & POLANI, P. E. Differential transmission of Down's syndrome (mongolism) through male and female translocation carriers. *Lancet*, 1961, 2, 956-958.
- HAMERTON, J. L., & STEINBERG, A. C. Progeny of D/G translocation heterozygotes in familial Down's syndrome. *Lancet*, 1962, 1, 1408.
- HANHART, E. Mongoloide Idiotie bei Mutter und zwei Kindern aus Inzesten. [Mongoloid idiocy in mother and two children resulting from incest.] *Acta Genet. med. Gemellol., Rome*, 1960, 9, 112-130.
- HANHART, E., DELHANTY, J. D. A., & PENROSE, L. S. Trisomy in mother and child. *Lancet*, 1961, 1, 403.
- HARNDEN, D. G., MILLER, O. J., & PENROSE, L. S. The Klinefelter-mongolism type of double aneuploidy. *Ann. hum. Genet.*, 1960, 24, 165-169.
- HAYASHI, T., HSU, T. C., & CHAO, D. A case

- of mosaicism in mongolism. *Lancet*, 1962, 1, 218-219.
- HIRSCH, J., & ERLIENMEYER-KIMLING, L. Behavior genetics. In H. Helson & W. Bevan (Ed.), *Theories and data in psychology*. New York: Van Nostrand, in press.
- ILBERY, P. L. T., LEE, C. W. G., & WINN, S. M. Incomplete trisomy in a mongoloid child exhibiting minimal stigmata. *Med. J. Aust.*, 1961, 48, 182-184.
- INGALLS, T. H. Pathogenesis of mongolism. *Amer. J. Dis. Children*, 1947, 73, 279-292.
- JACOBS, P. A., BAIKIE, A. G., COURT BROWN, W. M., & STRONG, J. A. The somatic chromosomes in mongolism. *Lancet*, 1959, 1, 710.
- JARVIK, L. F. Senescence and chromosomal changes. *Lancet*, 1963, 1, 114-115.
- JÉRÔME, H., LEJEUNE, J., & TURPIN, R. Étude de l'excrétion urinaire de certains métabolites du tryptophane chez les enfants mongoliens. [Study of urinary excretion of certain metabolites of tryptophane in mongoloid infants.] *CR Acad. Sci., Paris*, 1960, 251, 474-476.
- KALLMANN, F. J. New goals and perspectives in human genetics. *Acta Genet. med. Gemellol., Rome*, 1961, 10, 377-388.
- KALLMANN, F. J., & RAINER, J. D. Psychotherapeutically oriented counseling techniques in the setting of a medical genetics department. In B. Stokvis (Ed.), *Topical problems in psychotherapy*. Vol. 4. New York: Karger, 1963.
- KING, M. J., GILLIS, E. M., & BAIKIE, A. G. Alkaline-phosphatase activity of polymorphs in mongolism. *Lancet*, 1962, 2, 1302-1305.
- KRIVIT, W., & GOOD, R. A. Simultaneous occurrence of mongolism and leukemia. *AMA J. Dis. Children*, 1957, 94, 289-293.
- LANMAN, J. T., SKLARIN, B. S., COOPER, H. L., & HIRSCHHORN, K. Klinefelter's syndrome in a ten-month-old mongolian idiot: Report of a case with chromosomal analysis. *New Engl. J. Med.*, 1960, 263, 887-890.
- LEHMANN, O., & FORSSMAN, H. A. Klinefelter's syndrome and mongolism in the same person. *Acta paediat.*, 1960, 49, 536-539.
- LEHMANN, O., & FORSSMAN, H. A. Familial mongolism with chromosomal translocation also observed in normal boy member. *Acta paediat.*, 1962, 51, 6-12.
- LEJEUNE, J., GAUTIER, M., & TURPIN, R. Les chromosomes humains en culture de tissus. [Human chromosomes in tissue culture.] *CR Acad. Sci., Paris*, 1959, 248, 602-603. (a)
- LEJEUNE, J., GAUTIER, M., & TURPIN, R. Étude des chromosomes somatiques de neuf enfants mongoliens. [Study of somatic chromosomes of nine mongoloid infants.] *CR Acad. Sci., Paris*, 1959, 248, 1721-1722. (b)
- LELONG, M., BORNICHE, P., KREISLER, & BAUDY. Mongolien issu de mere mongolienne. [Mongoloid offspring of mongoloid mother.] *Arch. Franc. pediat.*, 1949, 6, 231-238.
- LEVAN, A., & HSU, T. C. The chromosomes of a mongoloid female, the mother of a normal boy. *Hereditas*, 1960, 46, 770-772.
- LINDSTEN, J., ALVIN, A., GUSTAVSON, K.-H., & FRACCARO, M. Chromosomal mosaic in a girl with some features of mongolism. *Cytogenetics*, 1962, 1, 20-31.
- LUBS, H. A. Causes of familial mongolism. *Lancet*, 1961, 2, 881.
- MACINTYRE, M. N., STAPLES, W. I., STEINBERG, A. G., & HEMPLE, J. M. Familial mongolism (trisomy-21 syndrome) resulting from a "15/21" chromosome translocation in more than three generations of a large kindred. *Amer. J. hum. Genet.*, 1962, 14, 335-344.
- McKUSICK, V. A. On the X chromosome of man. *Quart. Rev. Biol.*, 1962, 37, 69-176.
- MACLEAN, N., HARNDEN, D. G., & COURT BROWN, W. M. Abnormalities of sex chromosome constitution in newborn babies. *Lancet*, 1961, 2, 406-408.
- MAKINO, S., HONDA, T., & MITANI, M. Chromosome studies in children with Down's syndrome and their parents. *Proc. Jap. Acad.*, 1962, 38, 47-51.
- MENDEL, G. Versuche über Pflanzen-Hybriden. (Orig. publ. 1866) [Experiments in plant-hybridization.] In J. Peters (Ed.), *Classic papers in genetics*. Englewood Cliffs, N. J.: Prentice-Hall, 1959.
- MILLER, O. J., BREG, W. R., SCHMICKEL, R. D., & TRETER, W. A family with XXXXY male, a leukemic male, and two 21-trisomic mongoloid females. *Lancet*, 1961, 2, 78-79.
- MILLER, O. J., MITTWOCH, U., & PENROSE, L. S. Spermatogenesis in man, with especial reference to aneuploidy. *Heredity*, 1960, 14, 456.
- MOORE, K. L. Sex reversal in newborn babies. *Lancet*, 1959, 1, 217-219.
- MULLINS, D. H., ESTRADA, W. R., & GREASY, T. G. Pregnancy in an adult mongoloid female. *Obstet. Gynecol.*, 1960, 15, 781-783.
- NICHOLS, W. W., CORELL, L. L., FABRIZIO, D. P. A., BISHOP, H. C., & BOGGS, T. R., JR. Mongolism with mosaic chromosome pattern. *J. Pediat.*, 1962, 60, 69-76.
- O'BRIEN, D. G., & GROSHEK, A. The abnormality of tryptophane metabolism in children with mongolism. *Arch. Dis. Childh.*, 1962, 37, 17-20.
- ØSTER, J. *Mongolism* (op. ex domo biol. hered. hum. Univer. Hafniensis, Vol. 32).

- Copenhagen: E. Munksgaard, 1953.
- ØSTER, J. The causes of mongolism. *Danish med. Bull.*, 1956, 3, 158-164.
- PENROSE, L. S. Relative aetiological importance of birth order and maternal age in mongolism. *Proc. Roy. Soc. Lond. Ser. B*, 1934, 115, 431-450.
- PENROSE, L. S. The distal triradius on the hands of parents and sibs of mongol imbeciles. *Ann. hum. Genet.*, 1954, 19, 10-38. (a)
- PENROSE, L. S. Observations on the etiology of mongolism. *Lancet*, 1954, 2, 505-509. (b)
- PENROSE, L. S. Mongolism. *Brit. med. Bull.*, 1961, 17, 184-189.
- PENROSE, L. S. Paternal age in mongolism. *Lancet*, 1962, 1, 1101.
- PENROSE, L. S., ELLIS, J. R., & DELHANTY, J. D. R. Chromosomal translocation in mongolism and in normal relatives. *Lancet*, 1960, 2, 409-410.
- POLANI, P. E., BRIGGS, J. H., FORD, C. E., CLARKE, C. M., & BERG, J. M. A mongol girl with 46 chromosomes. *Lancet*, 1960, 1, 721-724.
- REHN, A. T., & THOMAS, E., JR. Family history of a mongolian girl who bore a mongolian child. *Amer. J. ment. Defic.*, 1957, 62, 496-499.
- RICHARDS, B. W., & STEWART, A. Mosaicism in a mongol. *Lancet*, 1962, 1, 275-276.
- SAWYER, G. M. Case report: Reproduction in a mongoloid. *Amer. J. ment. Defic.*, 1949, 54, 204-206.
- SAWYER, G. M., & SHAFTER, A. J. Reproduction in a mongoloid: A follow-up. *Amer. J. ment. Defic.*, 1957, 61, 793-795.
- SCHERZ, R. G. Negro female mongol with 46 chromosomes. *Lancet*, 1962, 2, 882.
- SCHLAUG, R. A mongolian mother and her child: A case report. *Acta genet.*, 1957, 7, 533-540.
- SHAW, M. W. Segregation ratios and linkage studies in a family with six translocation mongols. *Lancet*, 1962, 1, 1407.
- SHUTTLEWORTH, G. E. *Mentally-deficient children*. London: Lewis, 1895.
- SHUTTLEWORTH, G. E. Mongolian imbecility. *Brit. med. J.*, 1909, 2, 661-665.
- STILES, K. A., & GOODMAN, H. O. Reproduction in a mongoloid. *Acta Genet. med. Gemellol., Rome*, 1961, 10, 457-469.
- THULINE, H. C., & PRIEST, J. H. Pregnancy in a 14-year-old mongoloid. *Lancet*, 1961, 1, 1115-1116.
- TJIO, J. H., & LEVAN, A. The chromosome number of man. *Hereditas*, 1956, 42, 1-6.
- TJIO, J. H., & PUCK, T. T. Genetics of somatic mammalian cells: II. Chromosomal constitution of cells in tissue culture. *J. exp. Med.*, 1958, 108, 259-268. (a)
- TJIO, J. H., & PUCK, T. T. The somatic chromosomes of man. *Proc. Nat. Acad. Sci.*, 1958, 44, 1229-1237. (b)
- TOUGH, I. M., COURT BROWN, W. M., BAIKIE, A. G., BUCKTON, K. E., HARNDEN, D. G., JACOBS, P. A., KING, M. J., & MCBRIDE, J. A. Cytogenetic studies in chronic myeloid leukemia and acute leukemia associated with mongolism. *Lancet*, 1961, 1, 411-417.
- TURPIN, R., & CARATZALI, A. Conclusions d'une étude génétique de la langue plicaturée. [Conclusions of a genetic study of furrowed tongue.] *CR Acad. Sci., Paris*, 1933, 196, 2040-2042.
- TURPIN, R., & LEJEUNE, J. Chromosome translocations in man. *Lancet*, 1961, 1, 616-617.
- TURPIN, R., LEJEUNE, J., LAFOURCADE, J., & GAUTIER, M. Aberrations chromosomiques et maladies humaines: La polydyspondylie à 45 chromosomes. [Chromosomal aberrations and human maladies: Polydyspondily with 45 chromosomes.] *CR Acad. Sci., Paris*, 1959, 248, 3636-3638.
- WARKANY, J. Etiology of mongolism. *J. Pediatr.*, 1960, 56, 412-419.
- WATSON, J. D., & CRICK, F. H. C. The structure of DNA. *Cold Spr. Harb. Symp. quant. Biol.*, 1953, 18, 123-131.

(Received March 4, 1963)

## NOTE ON "PERSONALITY VARIABLES AND RESPONSE TO COLOR"

IRA M. STEISEL

*St. Christopher's Hospital for Children and Temple University Medical School*

Some criticisms of an article by Cerbus and Nichols (1963) relating to their report of author's work.

In their recent review of the literature on the relationship between the response to color and variables of personality, Cerbus and Nichols (1963) had occasion to refer to an earlier study of mine (Steisel, 1952) which attempted to relate some scores of the Rorschach test and measures of suggestibility. In their report there were a number of errors (both of omission and commission) which, perhaps, warrant correction so that the mistakes will not persist even though the overall conclusions of their review might be unaffected.

In the first place, they (Cerbus & Nichols, 1963) state that a confederate was used in an effort to influence the estimates of the amount of movement in the autokinetic situation (p. 569). No confederate was involved; rather the experimenter told the subject the "true" distance (actually a prearranged one).

Secondly, there were additional measures of suggestibility used in the original study beyond the change of estimate of movement as influenced by the experimenter's statements. Hull's Body-Sway test (1933) and the Ink-Blot Suggestion test of Eysenck and Furneaux (1945) were employed. In addition to these, a latency measure and the mean amount of movement reported (prior to subjects being given a suggestion of "correct" distance) were derived from the autokinetic procedure.

Despite their comment (Cerbus & Nichols, 1963, p. 569), *Sum C* was not utilized; although there were three other

measures of color responsivity obtained from the Rorschach: the absolute and the percentage of responses scored as *CF* as well as the average color score which had been proposed by Williams (1947). Nine additional Rorschach measures were reported.

Perhaps what is most distressing is that the two correlations (of a total of 72 reported) were singled out for discussion by Cerbus and Nichols (1963, p. 569). It is here contended that selecting them is a questionable procedure in light of the original finding that a total of only three correlations were greater than zero at the 5% level of confidence or beyond. The inescapable conclusion from the study, alas, is that no relationship exists between the Rorschach scores and the measures of suggestibility employed.

### REFERENCES

- CERBUS, C. & NICHOLS, R. C. Personality variables and response to color. *Psychol. Bull.*, 1963, 60, 566-575.
- EYSENCK, H. J., & FURNEUX, W. D. Primary and secondary suggestibility: An experimental and statistical study. *J. exp. Psychol.*, 1945, 35, 485-503.
- HULL, C. L. *Hypnosis and suggestibility*. New York: Appleton-Century, 1933.
- STEISEL, I. M. The Rorschach test and suggestibility. *J. abnorm. soc. Psychol.*, 1952, 47, 607-614.
- WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, 11, 21-29.

(Received December 9, 1963)

## SIGNIFICANCE LEVELS OF INDIVIDUAL CORRELATIONS SELECTED FROM MULTIVARIATE STUDIES

GEORGE CERBUS

*Mental Health Center, Marshalltown, Iowa*

AND ROBERT C. NICHOLS

*National Merit Scholarship Corporation, Evanston, Illinois*

Reply to criticisms of authors' report of Steisel's (1952) work.

Steisel (1964) has elaborated on our review's scant coverage of his study (Steisel, 1952), which was due to the necessarily selective nature of a review and to our evaluation of his original report. As Steisel implied, his clarifications do not change our interpretation of his findings.

Steisel's criticism of our selective reporting of his findings shows a misunderstanding of probability levels, which may or may not be affected when individual correlations are selected from a multivariate study depending on the method of selection. When one tests an hypothesis he is implicitly selecting a single relationship for study from an infinite number which could be computed. We focused on two of Steisel's correlations which were pertinent to the point being discussed and which provided partial replication of another study. Since these results were selected because of the nature of the variables, not the size of the correlations, the number of correlations

computed between the other (irrelevant) variables in the same study does not affect the significance level of the selected correlations.

We would like to second Cattell's (1958) recommendation that authors of reviews scan multivariate experiments for variables pertinent to the topics reviewed (as was done with Steisel's study). As long as correlations are selected because of the nature of the variables involved and not the size of the relationship, the usual probability levels apply.

### REFERENCES

- CATTELL, R. B. A need for alertness to multivariate experimental findings in integrative surveys. *Psychol. Bull.*, 1958, 55, 253-256.
- STEISEL, I. M. The Rorschach test and suggestibility. *J. abnorm. soc. Psychol.*, 1952, 47, 607-614.
- STEISEL, I. M. Note on "Personality variables and response to color." *Psychol. Bull.*, 1964, 61, 399.

(Received January 20, 1964)

# Psychological Bulletin

## ZERO CORRELATIONS AMONG TESTS OF INTELLECTUAL ABILITIES

J. P. GUILFORD

*University of Southern California*

It is often asserted that intercorrelations among tests of intellectual abilities are almost always positive, justifying the belief in a "g" factor and supporting the choice for oblique rotations of axes in factor analyses. Empirical information is presented from correlation matrices obtained in 13 typical analyses of intellectual measures showing that among more than 7000 correlation coefficients, 17-24% can be considered to be 0. It is estimated that under optimal conditions, when there are 15 factors, as few as 11% would need to be 0 to determine a clear, orthogonal simple structure.

Of a number of statements in the psychological literature that achieve the status of clichés, or nearly so, the assertion that practically all tests of intellectual abilities correlate positively is one of the most common. The statement, when it appears, can also be taken to imply significantly nonzero correlations. It is commonly used as "proof" that there is a universal or *g* factor on the one hand, and as logical support for the practice of oblique rotational solutions in factor analysis on the other. In discussing these points in a previous article (Guilford & Zimmerman, 1963), it was asserted that "hundreds of them (zero correlations) can be exhibited." It is the purpose of this note to support that statement by citing some empirical information.

Some years ago, the quoted statement could not be so confidently made. In the earlier efforts directed toward the analysis of intellectual abilities into their unique components, when most of the present-day knowledge of intellectual factors and sophistication in applications of factor analysis were unavailable,

many of the analyzed tests did intercorrelate substantially, or at least positively. From what is known now, there were essentially two reasons for those higher correlations. One reason was that the conception of intelligence was a relatively limited one, consequently, tests designed for this area of aptitude were less varied and more restricted to certain kinds that had much overlapping of factorial content. The other main reason was that the tests were factorially more complex; consequently the chances of one test having something in common with another in terms of common-factor variance were increased.

On the basis of present-day information, we can recognize the kinds of tests that are required to represent factors more univocally. It is easier to keep the factorial complexity of tests at a lower level, although complete success in this respect is still far from perfect. It is a matter of writing tests in such ways as to achieve the necessary experimental controls that are needed to focus examinees' efforts and methods in the favored directions. The development of the struc-

ture-of-intellect model, with its pointed, empirical definitions of intellectual abilities, demonstrated or still to be demonstrated, has helped considerably in guiding test construction in more definite directions. The very great broadening of the conception of the great domain of human intellect has also brought into consideration varieties of abilities not dreamed of before, thus leading to greater varieties of tests.

#### A SURVEY OF SOME CORRELATION MATRICES

The data used here to demonstrate the incidence of zero intercorrelations among tests of intellectual abilities were derived from a number of factor-analytic studies conducted in the Aptitudes Project at the University of Southern California during the past 15 years. Such correlations may be regarded as typical of those arising from the use of factor analysis in the Thurstone tradition. By this is meant that the interest is in discovering underlying psychological variables that can be regarded as ways of accounting for goodness of performances in different kinds of intellectual effort in the form of psychological tests.

Before the analysis, considerable thought is given to the number and nature of unique abilities involved within a circumscribed area of functioning, deriving hypotheses concerning what kinds of abilities will be demonstrated by the analysis. Tests are developed with an effort to measure each hypothesized ability and that ability only. Each test is designed to measure something very different from the basic abilities measured by tests designed for other basic abilities. Consequently, there is a frank effort to achieve minimal correlation between every pair of tests of two different factors. If the two factors are completely uncorrelated in the population of examinees, the goal of a zero

correlation is a realistic one. Any correlation greater than zero between two factors limits the possibility of achieving a zero correlation between the pairs of tests for those two factors, so long as the tests have nonzero reliability.

Thus, one may say that, under the circumstances of the studies represented in the data, conditions were about as favorable as it has been possible to achieve, during the past 15 years. Further gains in sophistication might permit a more general improvement in reduction of intercorrelations where factors are not in common to tests. On the other hand, a more random selection of tests for factor-analytic batteries could certainly be expected to reduce the proportions of zero or near-zero correlations. The latter conditions would hardly be tolerated by those who wish to make progress in isolating new factors.

Considering all 23 analyses of intellectual abilities that have been performed by the Aptitudes Project, 19 of which have been reported in the literature, 13 were selected for use in the survey of this note. Selection was made without regard to inspection of the correlation matrices. It aimed to use the more representative analyses, in terms of the kinds of test variables employed. For example, matrices were rejected where more than two variables were nonintellectual (such as a test for finger speed or the variable of sex membership), with one exception of selection of one study that had three such variables. Where one or more such variables were involved, they were excluded in the tally of correlation coefficients, for they provided biased numbers of near-zero correlations. Variables were also excluded whose communalities were less than .30, since low communality puts a limitation on the size of correlation coefficients. Such instances were extremely rare in the matrices used. In three analyses

(Report Number 26), one variable (Reading Comprehension) was undoubtedly a factorially complex score, which might have increased its chances of nonzero correlations. But most of the other variables were tests of divergent-production abilities, which usually correlate low, even zero, with tests of the traditional types of IQ tests, such as Reading Comprehension.

This kind of selection left the use of correlation matrices from *Reports from the Psychological Laboratory*, Numbers 3, 8, 9, 16, 17, 18, 23, and 26.<sup>1</sup> There was one matrix each from most of these reports, there were three from Report Number 16, and four from Report Number 26. Most of the analyses were based upon subjects who were males with generally higher-than-average IQ levels, near the age of 20, in military training that ordinarily led to officer commissions. In Report Number 26 the subjects were ninth-grade students in two southern California high schools. All samples exceeded 200 in number, with an  $N$  of about 225 being rather typical. The coefficients were usually Pearson  $r$ 's, or estimates of Pearson  $r$ 's from point biserials or cosine-pi estimates from  $2 \times 2$  contingency tables.<sup>2</sup>

## Results

The frequency distribution of all coefficients of correlation from the 13 matrices is presented in Table 1, with proportions of the coefficients within each class interval and also cumulative proportions. The total range is from three  $r$ 's that fell below  $-.2$  to a single  $r$  that was above  $+.7$ . If we define a zero  $r$  as being within the range from  $-.1$  to  $+.1$ , we see that  $.1357 + .0357 = .1711$  of the coefficients are essentially

<sup>1</sup> See Michael, Comrey, and Fruchter (1962) for a listing of all these reports.

<sup>2</sup> I am indebted to Gail O'Connor for material assistance in compiling this statistical information.

TABLE 1  
SUMMARY STATISTICS ON DISTRIBUTIONS OF  
COEFFICIENTS OF CORRELATION IN 13  
FACTOR ANALYSES OF TESTS OF  
INTELLECTUAL ABILITIES

Class intervals of $r$	$f$	$p$	$cp$
.7 to .799	1	.0001	.9999
.6 to .699	18	.0025	.9998
.5 to .599	121	.0171	.9973
.4 to .499	561	.0792	.9802
.3 to .399	1373	.1939	.9010
.2 to .299	2005	.2831	.7071
.1 to .199	1743	.2461	.4240
.0 to .099	959	.1354	.1779
-.1 to -.001	253	.0357	.0425
-.2 to -.101	45	.0064	.0068
-.3 to -.201	3	.0004	.0004
Sum	1602.50	.7082	.9999
Mean	.226		

zero in value, or about 17%, or one sixth. Below the level of  $-.1$  are 48 coefficients, or about two thirds of 1% of the  $r$ 's, which function like zero correlations in indicating independence of factors and suggesting orthogonal rotations of axes. The percentage of correlations below  $+.1$  varies a great deal from one analysis to another, the range being from 5% to 34%, where the percentage over all samples is approximately 18.

The choice of  $r$ 's below  $+.1$  in defining a zero coefficient is an arbitrary one. Let us consider a rational guide, in the form of an estimate of sampling errors and statistical inference. Considering that the typical size of sample is 225, as mentioned earlier, the standard error of  $r$ , when the population value of  $r = .0$ , is .067. A departure from zero of two standard errors is then .134, not .100. The null hypothesis cannot be rejected at the .05 level of confidence with obtained  $r$ 's within the range from  $-.134$  to  $+.134$ , which would include a substantially higher proportion than the

17% between  $-1$  and  $+1$ ; possibly as many cases as 24%. This estimation is admittedly very rough, for some experimental samples of examinees contained far more than 225 and none was smaller than 200.

#### SOME FURTHER CONSIDERATIONS

From this rather extensive information of a typical nature, the evidence is quite clear that zero correlations are quite common among tests of intellectual abilities. The general level of such coefficients is indicated by a mean  $r$  of .226.<sup>3</sup> The distribution of  $r$ 's about that central value is nearly symmetrical, with a bit of positive skewing. There are a number of conditions that should lead us to expect the form that such a distribution should take.

Let us consider the special case of a perfect, orthogonal, simple structure; that is, each test loaded on only one factor and the factors uncorrelated. Thus, if there were  $m$  factors and  $k$  tests per factor, there should be  $mk(k-1)/2$  positive correlations and  $k^2m(m-1)/2$  zero correlations. To take a somewhat typical, realistic, case with 15 factors and 3 tests per factor, the number of positive correlations would be 45 and the number of zero correlations 945; or a total of 990, which should equal  $km(km-1)/2$ . In this special case, about 95% of the  $r$ 's would be zero, which is far above the 17% or

even the 24% that were found empirically.

But such extreme or ideal instances are never found, nor should they be expected; nor are they needed in achieving an orthogonal simple-structure solution. With graphic rotations of axes, at least, so long as there is one zero correlation between one pair of tests, each representing one of a pair of the factors, a clear, orthogonal solution can be achieved. This means, of course, that the much smaller number of zero correlations would have to be optimally distributed, but it would not be limited to any one particular distribution, among pairs of tests. The number of such zero correlations can be surprisingly small. This number is given by the quantity  $m(m-1)/2$ , which, in the case of 15 factors would be 105. Out of the total of 990  $r$ 's this would be approximately 11%, which is distinctly below the empirical 17-24%. Where the zero  $r$ 's are not located according to one of the optimal distributions among the pairs, of course, there would be departure from a complete, orthogonal solution. On the other hand, psychologically satisfying solutions have been obtained with less than 11% of the correlations in the category below  $+1$ .

#### REFERENCES

- GUILFORD, J. P., & ZIMMERMAN, W. S. Some variable-sampling problems of axes in factor analysis. *Psychol. Bull.*, 1963, 60, 289-301.  
MICHAEL, W. B., COMREY, A. L., & FRUCHTER, B. J. P. Guilford: Psychologist and teacher. *Psychol. Bull.*, 1962, 59, 1-34.

<sup>3</sup> Averaging the corresponding Fisher  $z$  values gave a very similar result.

(Received May 23, 1963)

# EXPERIMENTER OUTCOME-ORIENTATION AND THE RESULTS OF THE PSYCHOLOGICAL EXPERIMENT<sup>1</sup>

ROBERT ROSENTHAL  
*Harvard University*

This review considers evidence bearing on the question of whether an E's orientation (expectations and wishes) toward the results of his research may partially determine those results. The evidence is drawn from everyday life, medicine, clinical psychology, survey research, and experimental behavioral research.

In any science, experimenters have some orientation towards the outcome or results of their research. Rarely is this orientation one of truly dispassionate disinterest. Variables are not chosen for inclusion in research by using tables of random numbers. They are, rather, chosen because the experimenter has certain expectations about the relationship or lack of relationship between the selected variables and certain other variables. A superficial exception to this might be seen in so-called heuristic hunts for relationships, which are perhaps more common to the behavioral sciences. Even here, however, the inclusion of variables is not on a random basis, and certain relationships appear more likely to be found than others.

Experimenters then often, if not always, have some sort of expectations about how the data will fall. Also often, if not always, they care about how these data fall. Some outcomes may be expected more than others; some outcomes may be desired more than others. Our purpose here is to discuss the question of whether experimenter's orientation (expectations and wishes) can affect the data actually obtained in his research. We are not so much concerned here with the problem of choice of experimental

design or procedure and the fact that certain designs and procedures may unintentionally be more or less favorable to obtaining expected or unexpected data. Neither are we concerned with the problems of statistical tests of hypotheses and the fact that uniquely most powerful statistics may unintentionally be employed when the expectation is to be able to reject the null hypothesis, while less powerful statistics may be employed when the expectation is to be unable to reject the null hypothesis. These are interesting questions but will not be considered here. Our usage of "results" or "outcome" will be restricted to the raw data obtained by experimenters from their subjects.

The effects of experimenter's outcome orientation, or bias, were seriously considered by Wilson (1952), the physical scientist. Wilson felt that positive or expected data might too often occur because of researchers' interest in the outcome of their experiments. Their expectancies about data might determine in part the data obtained. This notion seems most related to Merton's (1948) concept of self-fulfilling prophecy. One prophesies an event and the expectation of the event, then changes the prophet's behavior in such a way as to make the predicted event more likely. Related, too, is Heider's (1958) concept of "personal causality" and his dis-

<sup>1</sup> Preparation of this paper was facilitated by a grant, G-24826, from the Division of Social Sciences of the National Science Foundation.

cussion of the fulfillment of personal expectancies.

#### OUTCOME-ORIENTATION EFFECTS IN EVERYDAY LIFE

The way a man golfs or bowls may be determined by his expectancy of his performance. Of greater interest here is the notion that how one person expects another to perform these activities can determine in part how he actually does perform. In an intensive study of a social group of young men, Whyte (1943) found that the group, and especially its leaders, "knew how well a man should bowl." This "knowledge" or expectancy probably did partially determine that man's actual performance. Perhaps the morale-building banter offered the one expected to do well helped him to do well. Similarly, the communicated expectancy to another that he would do poorly "shook him up" sufficiently to interfere with his performance (see also Back, 1951).

Fascinating data collected at the Bank Street College of Education suggest that in the schoolroom as in the bowling lanes, expectancies may be powerful forces determining others' behavior. Data described by John Niemeyer,<sup>2</sup> President of the College, lend support to the hypothesis that lower-class, minority-group children are low achievers, at least partly because of their teachers' expectation that these students are really not educable.

#### OUTCOME-ORIENTATION EFFECTS IN CLINICAL PRACTICE

As highly skilled a clinician as Fromm-Reichmann (1950) was impressed by the effects of the self-fulfilling prophecy, although she did not use that term. She spoke rather, as other clinicians have, of iatrogenic psychiatric incurabilities. The therapist's expectancy, she felt, might

determine whether given symptoms might be relieved or cured. This clinical impression is somewhat supported by the work of Heine and Trosman (1960) who felt that the variable significant for a patient's continuance in psychotherapy was that of mutuality of expectation between therapist and patient. Goldstein (1960) found no client-perceived personality change due to psychotherapy related to therapist's expectancy of such change. However, therapist's expectancy was related to duration of psychotherapy. Additionally, Heller and Goldstein (1961) found therapist's expectation of client improvement significantly correlated (.62) with change in client's attraction to therapist. These workers also found that after 15 sessions, clients' behavior was no more independent than before, but their self-descriptions were of more independent behavior. The therapists generally were favorable to increased independence and tended to expect successful cases to show this decrease in dependency. Clients may well have learned from their therapists that independent-sounding verbalizations were desired and thereby served to fulfill their therapist's expectancy. The role of expectancy in the psychotherapeutic situation has been most fully discussed and reviewed by Goldstein (1962).

But psychotherapy is not the only realm of clinical practice in which expectancy effects may determine outcomes. The fatality rates of delirium tremens have recently not exceeded about 15%. However, from time to time new treatments of greatly varying sorts are reported to reduce this figure almost to zero. Gunne's (1958) work in Sweden summarized by the *Quarterly Journal of Studies on Alcohol* Editorial Staff (1959) showed that any change in therapy led to a drop in mortality rate. One interpretation of this finding is that

<sup>2</sup> Personal communication, 1961.

the innovator of the new treatment expects a decrease in mortality rate, an expectancy which leads to subtle differential patient care over and above the specific treatment under investigation. A prophesy again may have been self-fulfilled.

In the practice of medicine in general, the role of physician expectancy looms large. In a very comprehensive paper dealing with placebo effects, Shapiro (1960) cites the well-known admonition: "You should treat as many patients as possible with the new drugs while they still have the power to heal [p. 114]." The wisdom of this statement may have its basis in the concept of the physician's faith in the power of the drug. This "faith" may have at its core expectancy as we are discussing it. The physician's expectancy about the efficacy of a treatment may be subtly communicated to the patient with resulting influence on the patient's psychobiological response.

#### OUTCOME-ORIENTATION EFFECTS IN SURVEY RESEARCH

Perhaps the classic work in this area was that of Rice (1929). A sample of 2,000 applicants for charity was interviewed by a group of 12 skilled interviewers. Interviewers talked individually with their respondents who had been assigned in a wholly nonselected manner. Respondents ascribed their dependent status to factors predictable from a knowledge of the interviewers' outcome orientations. Thus, one of the interviewers who was a staunch prohibitionist obtained three times as many responses blaming alcohol as did another interviewer regarded as a socialist, who in turn obtained half again as many responses blaming industrial factors as did the prohibitionist interviewer. Rice concluded that the outcome orientation or bias of the interviewer was somehow communicated to the respondent who

then replied as expected. Hyman, Cobb, Feldman, Hart, and Stember (1954) took vigorous exception to Rice's interpretation and preferred to ascribe his remarkable results to errors of recording or interpretation. The plain fact, of course, is that we cannot say whether these effects were ones which actually changed respondents' replies or not. There is no question in either case that the results of the research were affected by the investigators' outcome orientation.

One of the earliest studies deliberately creating differential expectancies in interviewers was that conducted by Harvey (1938). Each of six boys was interviewed by each of five young post-graduates. The boys were to report to the interviewers on a story they had been given to read. Interviewers were to use these reports to form impressions of the boys' character. Each interviewer was given some contrived information about the boys' reliability, sociability, and stability, but told not to regard these data in assessing the boys. Standardized questions asked of the interviewers at the conclusion of the study suggested that biases of assessment occurred even without interviewers' awareness and despite conscious resistance to bias. Harvey felt that the interviewers' bias evoked a certain attitude towards the boys which in turn determined the behavior to be expected and then the interpretation given. Note how neatly this formulation fits the model put forth by Merton. Again, we cannot be sure that subjects' responses were actually altered by interviewer expectancies. The possibility, however, is too provocative to overlook.

More recent evidence for an expectancy (outcome orientation) bias comes from the work of Hanson and Marks (1958). The most thorough discussion of this problem for the survey research

literature is that by Hyman et al. (1954), which also carries an extensive bibliography.

#### OUTCOME-ORIENTATION EFFECTS IN EXPERIMENTAL RESEARCH

It is well known that a great many studies have been conducted to establish the validity or invalidity of the Rorschach technique of personality assessment. A systematic study of 168 of these studies was undertaken by Levy and Orr (1959) who categorized each study on each of the following dimensions: the academic versus nonacademic affiliation of the author, whether the study was designed to assess construct versus criterion validity, and whether the outcome of the study was favorable or unfavorable to the hypothesis of Rorschach validity. Results showed that academicians were more interested in construct validity and that their outcomes were relatively more favorable to construct validation and less favorable to criterion validation. On the basis of their findings, these workers called for more intensive study of the researcher himself. "For, intentionally or not, he seems to exercise greater control over human behavior than is generally thought [p. 83]." We cannot safely conclude that the findings reported were a case of the effect of outcome orientation or bias. It might have been that the choice of specific hypotheses for testing, or that the choice of manner of testing them determined the apparently biased outcomes. At the very least, however, this study accomplished its task of calling attention to potential biasing effects of experimenters.

Perhaps the earliest study which employed a straightforward experimental task and actually manipulated an outcome-orientation variable was that of Stanton and Baker (1942). In their study, 12 nonsense geometric figures

were presented to a group of 200 undergraduate subjects. After several days, retention of these figures was measured by five experienced workers. Experimenters were supplied with a key of "correct" responses, some of which were actually correct but some of which were incorrect. All experimenters were explicitly warned to guard against any bias associated with their having the keys before them and therefore influencing their subjects to guess correctly. Results showed that the experimenter obtained outcomes in accordance with his expectations. When the item on the key was correct, the subject's response was more likely to be correct than when the key was incorrect. In a careful replication of this study, Lindzey (1951) emphasized to his experimenters the importance of keeping the keys out of the subjects' view. This study failed to confirm the Stanton and Baker findings. The 85 subjects of Lindzey's study were much more of a volunteer population than were the subjects of the original study. We simply cannot say whether this fact might have accounted (in whole or in part) for the difference. Another replication by Friedman (1942) also failed to obtain the significance levels obtained in the original. Still, significant results of this sort, even occurring only in one out of three experiments, cannot be dismissed lightly. Stanton (1942, see pp. 16-17) himself presented further evidence which strengthened his conclusions. He employed a set of nonsense materials, 10 of which had been presented to subjects and 10 of which had not. Experimenters were divided into three groups. One group was correctly informed as to which 10 materials had been exposed, another group was incorrectly informed, while the third group was told nothing. The results of this study also indicated that the materials which experimenters expected to be

more often chosen were, in fact, more often chosen.

An experiment analogous to those just described was conducted in a psychophysical laboratory by workers (Warner & Raible, 1937) who interpreted their study within the framework of parapsychological phenomena. The study involved the judgment of weights by subjects who could not see their experimenter. The latter kept his lips tightly closed to prevent unconscious whispering (Kennedy, 1938). In half the experimental trials, the experimenter knew the correct response from a key. Of the 17 subjects, 6 showed a standard error of 1.0 or more from a 50-50 distribution of errors. All 6 of these subjects made fewer errors on trials on which the experimenter knew which weight was lighter or heavier. At least for those subjects who were somewhat affected by the experimenter's knowledge of the correct response, the authors' conclusion seems justified. As an alternative to the interpretation of these results as extrasensory perception (ESP) phenomena, they suggested the possibility of some form of auditory cue transmission to the subjects.

Among the most recent studies in the area of ESP are those by Schmeidler and McConnell (1958). These workers found that subjects who believed ESP possible ("sheep") performed better at ESP tasks than did subjects who did not believe ESP possible ("goats"). These workers suggested that an experimenter by his presentation, might affect subjects' self-classification, thereby increasing or decreasing the likelihood of successful ESP performance. Similarly, Anderson and White (1958) found that teachers' and students' attitudes toward each other might influence performance in classroom ESP experiments. The mechanism operating here might also have been one of certain teachers' ex-

pectancies which were communicated to the children whose self-classification as sheep or goats might thereby be affected. The role of the experimenter in the results of ESP research has been discussed by Crumbaugh (1959) as a source of evidence against the existence of the phenomenon. We file no brief here for or against ESP, but suggest that if, in carefully done experiments, certain types of experimenters obtain certain types of ESP performances in a predictable manner (as suggested by the studies cited), that further evidence for the effects of experimenter outcome-orientation will have been adduced (Rhine, 1959).

In a more traditional area of psychological research—memory—Ebbinghaus (1913) called attention to similar experimenter effects. In his own research he noted that his expectancy of what data he would obtain affected the data he subsequently did obtain. He pointed out, furthermore, that the experimenter's knowledge of this expectancy was not sufficient to control the phenomenon. This finding has been unfortunately neglected by many subsequent researchers in the area.

Another possible case has been described by Stevens (1961). He discussed the controversy between Fechner and Plateau over the results of bisection experiments to determine the nature of the function describing the operating characteristics of a sensory system. Plateau held that it was a power rather than a log function. Delboeuf carried out experiments for Plateau, but obtained data approximating the Fechnerian prediction of a log function. Stevens puzzled over these results which may be interpreted within the notion of experimenter outcome-orientation. Either by implicitly expecting the Fechnerian outcomes or by attempting to guard against an anti-Fechnerian bias, Del-

boeuf may have influenced the outcome of his studies.

It would appear that Pavlov was aware of the possibility that experimenter outcome-orientation might affect the results of experiments. In an exchange of letters in *Science*, Zirkle (1958) and Razran (1959) in discussing Pavlov's attitude toward the notion of the inheritance of acquired characteristics, gave credence to a statement by Gruenberg (1929):

In an informal statement made at the time of the Thirteenth International Physiological Congress, Boston, August, 1929, Pavlov explained that in checking up these experiments, it was found that the apparent improvement in the ability to learn, on the part of successive generations of mice, was really due to an improvement in the ability to teach, on the part of the experimenter! And so this "proof" of the transmission of modifications drops out of the picture, at least for the present [p. 327].

Wherry<sup>3</sup> has told of an experiment in which rats were able to discriminate colors, but only when the experimenter was in the room. Christie's (1951) interpretation of some differences between Iowa and Berkeley rats also suggests the possibility of experimenter effects associated with his outcome orientation.

But perhaps the best-known and most instructive case illustrating the effects of outcome orientation is that of Clever Hans (Pfungst, 1911). By means of tapping his hoof, the horse of von Osten was able to spell, read, and solve problems of arithmetic and musical harmony. Unlike the owners of other performing animals, Hans' owner did not profit from his animal's talent, and permitted any serious investigator to test Hans even in von Osten's absence. Pfungst, and his colleague Stumpf, undertook to discover the secret of Hans' talents.

A series of brilliant and painstaking experiments revealed that Hans' questioners cued him unintentionally. A for-

ward inclination of the questioner's head served as signal to Hans to begin his hoof tapping. A slight upward motion of the questioner's head or eyebrows served as signal for Hans to stop his tapping. Hans' amazing talents, then, may be viewed as an illustration of the power of the self-fulfilling prophecy. Questioners, even skeptical ones, expected Hans to know the correct answers to their queries. Their expectation was reflected in their signal to Hans that they awaited the cessation of his tapping. This signal brought on the expected cessation and Hans was correct again.

Pfungst aptly summarized the difficulties in uncovering the nature of Clever Hans' talents by speaking of "looking for, in the horse, what should have been sought in the man."

Turning to a more recent example of possible outcome-orientation effects, we will describe an experiment dealing with the Freudian defense mechanism of projection (Rosenthal, 1956). A total of 108 subjects was randomly divided into three groups each receiving success, failure, or neutral experience on a task structured as and simulating a standardized test of intelligence. Before the subjects' experimental-treatment condition was imposed, they were asked to rate the degree of success or failure of persons pictured in photographs. Immediately after the experimental manipulation, the subjects were asked to rate an equivalent set of photos on their degree of success or failure. The dependent variable was the magnitude of the difference scores from pre- to postratings of the photographs. It was hypothesized that the success-treatment condition would lead to greater subsequent perception of other people's success, while the failure-treatment condition would lead to greater subsequent perception of other people's failure as measured by the pre-post difference scores.

<sup>3</sup> Personal communication, 1960.

An analysis (which was essentially unnecessary to the main purpose of the study) was performed which compared the mean preratings of the three experimental-treatment conditions. Preratings by subjects in the success-treatment group were significantly lower and less extreme than the prerating by subjects in the other conditions. In terms of the hypothesis under test, a lower prerating by this group would tend to lead to significantly different difference scores if the postratings were similar for all treatment conditions. Without the investigator's awareness, the cards had been stacked in favor of obtaining results confirming the hypothesis under test. It should be emphasized that the success and failure groups' instructions had been verbally identical during the prerating phase of the experiment.

The investigator, however, was aware for each subject which experimental treatment the subject would subsequently be administered.

The implication is that in some subtle manner, perhaps by tone, or manner, or gestures, or general atmosphere, the experimenter, although formally treating the success and failure groups in an identical way, influenced the success Ss to make lower initial ratings and thus increase the experimenter's probability of verifying his hypothesis [Rosenthal, 1956, p. 44].

Reports of the findings of the sort just presented are not numerous and virtually never published. Nevertheless, their occurrence can be documented.<sup>4</sup> Allusions to the effects of the experimenter's outcome orientation<sup>5</sup> in general have been made by Edwards (1950); Feldman (1956); Foster (1923); Riecken (1962); Cohen, Silverman, Bressler, and Shmavonian (1961); and half facetiously by Ammons and Ammons (1957); and Rotter.<sup>6</sup>

<sup>4</sup> O. Gardebring, 1962; J. Gengerelli, 1956; G. Mount, 1956; and G. Rosenwald, 1963; personal communications.

<sup>5</sup> A series of experiments specifically de-

## REFERENCES

- AMMONS, R. B., & AMMONS, CAROL H. ESP and PK: A way out? *N. Dak. Quart.*, 1957, 25, 119-121.
- ANDERSON, MARGARET, & WHITE, RHEA. A survey of work on ESP and teacher-pupil attitudes. *J. Parapsychol.*, 1958, 22, 246-268.
- BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, 46, 9-23.
- CHRISTIE, R. Experimental naivete and experimental naivete. *Psychol. Bull.*, 1951, 48, 327-339.
- COHEN, S. I., SILVERMAN, A. J., BRESSLER, B., & SHMAVONIAN, B. Problems in isolation studies. In P. K. Solomon, P. E. Kubzansky, P. H. Leiderman, J. H. Mendelson, R. Trumbull, & D. Wexler (Eds.), *Sensory deprivation*. Cambridge: Harvard Univer. Press, 1961.
- CRUMBAUGH, J. D. ESP and flying saucers: A challenge to parapsychologists. *Amer. Psychologist*, 1959, 14, 604-606.
- EBBINGHAUS, H. *Memory: A contribution to experimental psychology*. (Orig. publ. 1885; trans. by H. A. Ruger & Clara E. Bussenius) New York: Teachers College, Columbia University, Bureau of Publications, 1913.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
- FELDMAN, P. E. The personal element in psychiatric research. *Amer. J. Psychiat.*, 1956, 113, 52-54.
- FOSTER, W. S. Experiments on rod-divining. *J. appl. Psychol.*, 1923, 7, 303-311.
- FRIEDMAN, PEARL. A second experiment on interviewer bias. *Sociometry*, 1942, 5, 378-379.
- FROMM-REICHMANN, FRIEDA. *Principles of intensive psychotherapy*. Chicago: Univer. Chicago Press, 1950.
- GOLDSTEIN, A. P. Therapist and client expectation of personality change in psychotherapy. *J. counsel. Psychol.*, 1960, 7, 180-184.
- GOLDSTEIN, A. P. *Therapist-patient expectations in psychotherapy*. New York: Pergamon Press, 1962.
- GRUENBERG, B. C. *The story of evolution*. New York: Van Nostrand, 1929.
- GUNNE, L. M. Mortaliteten vid delirium tremens. [Mortality in delirium tremens.] *Nord. Med.*, 1958, 60, 1021-1024.
- HANSON, R. H., & MARKS, E. S. Influence of

signed to investigate the occurrence and nature of the effects of the experimenter's outcome-orientation has recently been summarized elsewhere (Rosenthal, 1963).

<sup>6</sup> Personal communication, 1961.

- the interviewer on the accuracy of survey results. *J. Amer. Statist. Ass.*, 1958, 53, 635-655.
- HARVEY, S. M. A preliminary investigation of the interview. *Brit. J. Psychol.*, 1938, 28, 263-287.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- HEINE, R. W., & TROSMAN, H. Initial expectations of the doctor-patient interaction as a factor in continuance in psychotherapy. *Psychiatry*, 1960, 23, 275-278.
- HELLER, K., & GOLDSTEIN, A. P. Client dependency and therapist expectancy as relationship maintaining variables in psychotherapy. *J. consult. Psychol.*, 1961, 25, 371-375.
- HYMAN, H. H., COBB, W. J., FELDMAN, J. J., HART, C. W., & STEMBER, C. H. *Interviewing in social research*. Chicago: Univer. Chicago Press, 1954.
- KENNEDY, J. L. Experiments on "unconscious whispering." *Psychol. Bull.*, 1938, 35, 526. (Abstract)
- LEVY, L. H., & ORR, T. B. The social psychology of Rorschach validity research. *J. abnorm. soc. Psychol.*, 1959, 58, 79-83.
- LINDZEY, G. A note on interviewer bias. *J. appl. Psychol.*, 1951, 35, 182-184.
- MERTON, R. K. The self-fulfilling prophecy. *Antioch Rev.*, 1948, 8, 193-210.
- PFUNGST, O. *Clever Hans (the horse of Mr. von Osten): A contribution to experimental, animal, and human psychology*. (Trans. by C. L. Rahn) New York: Holt, 1911.
- QUARTERLY JOURNAL OF STUDIES ON ALCOHOL EDITORIAL STAFF. Mortality in delirium tremens. *N. Dak. Rev. Alcoholism*, 1959, 4, 3. (Abstract)
- RAZRAN, G. Pavlov the empiricist. *Science*, 1959, 130, 916.
- RHINE, J. B. How does one decide about ESP? *Amer. Psychologist*, 1959, 14, 606-608.
- RICE, S. A. Contagious bias in the interview: A methodological note. *Amer. J. Sociol.*, 1929, 35, 420-423.
- RIECKEN, H. W. A program for research on experiments in social psychology. In N. F. Washburne (Ed.), *Decisions, values and groups*. Vol. 2. New York: Pergamon Press, 1962. Pp. 25-41.
- ROSENTHAL, R. An attempt at the experimental induction of the defense mechanism of projection. Unpublished doctoral dissertation, University of California, Los Angeles, 1956.
- ROSENTHAL, R. On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *Amer. Scientist*, 1963, 51, 268-283.
- SCHMEIDLER, GERTRUDE, & MCCONNELL, R. A. *ESP and personality patterns*. New Haven: Yale Univer. Press, 1958.
- SHAPIRO, A. K. A contribution to a history of the placebo effect. *Behav. Sci.*, 1960, 5, 109-135.
- STANTON, F. Further contributions, twentieth anniversary of the Psychological Corporation and to honor its founder, James McKeen Cattell. *J. appl. Psychol.*, 1942, 26, 8-23.
- STANTON, F., & BAKER, K. H. Interviewer bias and the recall of incompletely learned materials. *Sociometry*, 1942, 5, 123-134.
- STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, 133, 80-86.
- WARNER, L., & RAIBLE, MILDRED. Telepathy in the psychophysical laboratory. *J. Parapsychol.*, 1937, 1, 44-51.
- WHYTE, W. F. *Street corner society*. Chicago: Univer. Chicago Press, 1943.
- WILSON, E. B. *An introduction to scientific research*. New York: McGraw-Hill, 1952.
- ZIRKLE, C. Pavlov's beliefs. *Science*, 1958, 128, 1476.

(Received February 18, 1963)

# ROLE OF MUSCLE TENSION IN PERSONALITY THEORY

IRIS BALSCHAN GOLDSTEIN<sup>1</sup>

*Psychosomatic and Psychiatric Institute  
Michael Reese Hospital, Chicago, Illinois*

The relationship of muscle tension to personality is reviewed through the 19th century to the present. Although the methods of measurement are often only indirectly related to one another certain trends emerge. Skeletal muscle tension tends to be high: in the hyperactive person, during emotional stress, in psychotic and neurotic groups, and particularly in disorders in which anxiety is the major concomitant. The mechanism of symptom specificity, whereby a particular muscle develops symptoms due to its constant state of sustained contraction, is also discussed.

The concept of muscle tension has been a very important one in the psychological literature. From the 1920s to the present it has aroused considerable speculation and stimulated a good deal of research. It has been related to learning (Bills, 1927; Courts, 1942), motivation (Bartoshuk, 1955), difficulty of mental work (Stroud, 1931), degree of arousal (Kennedy & Travis, 1948), reaction time (Teichner, 1957), attention (Wallerstein, 1954), and thinking (Jacobson, 1932), to mention a few. The body of this paper will be concerned with its role in personality theory.

In order to discuss muscle tension meaningfully it becomes necessary to clarify what one means by this term. Jacobson (1938), for example, refers to muscle tension in his book on *Progressive Relaxation* as a somewhat vaguely defined condition of nervous hypertension or nervous hyperexcitability. In some cases tension appears to be synonymous with tonus, while other writers attempt to separate the two terms and to confine tonus to a state of sustained contraction due to a continuous barrage of nerve impulses. Although Freeman (1948) speaks of the two as similar, he feels that tension is subject to cerebral influences.

<sup>1</sup> The author wishes to express her thanks to M. A. Wenger for his comments on the original draft of this paper.

Underlying the various descriptions of muscle tension one finds the idea of skeletal muscular contraction. For this reason, it appears to be most meaningful to define muscle tension in terms of muscular contraction rather than arguing over the different definitions. In states of complete relaxation there would be no muscular contraction and, therefore, no tension (Broman, 1949). When the muscle is stimulated, however, it contracts and there are simultaneous electrical, chemical, structural, and thermal changes.

A major difficulty in attempting to define muscle tension is that it frequently depends on what measure is used to obtain it. In an extensive article by Davis (1942), different measures that have been claimed to be indicators of muscle tension were reviewed. They range from modern electrical recording devices to some rather indirect means of assessment. It is important to keep in mind the method of measuring tension in each of the studies in this area.

## INDIVIDUAL DIFFERENCES IN MUSCLE TENSION

In the early twentieth century many writers speculated about the apparent relationship between the degree to which an individual expressed himself through the skeletal muscular system and such

traits as social extraversion, energy level, vitality, and excitability. While some of the early studies were based upon experimentation, a rather large number were merely clinical accounts of personality types. In addition, estimates of muscle tension were crude and frequently were based only on cursory descriptions of individuals.

There seems to be a rather general belief that the person who was constantly in a state of heightened tension would display this tension through certain expressive movements. On the basis of such a generalization, personality theorists attempted to show how a person's gait, posture, and gesture revealed relatively enduring personality attributes, independently of any mood or situational changes.

Muscle tension as revealed in voice and posture was believed by Dunbar (1946) and Deutsch (1947) to represent an individual's means of expressing habitual defenses. Particularly in states of conflict were defenses and repressed emotions reflected in expressive movements.

As far back as 1890, William James divided individuals into hyperactive and hypoactive types. Similar categorizations were postulated by Davenport (1915), Downey (1923), Allport and Vernon (1933), and Wolff (1933). At one end of the continuum was the calm, phlegmatic individual; while at the other extreme was the highly excitable person. These types were related by some (Hammett, 1920; Johnson, 1928) to the frequency and intensity of muscular contraction under emotional stimulation. McKinley and Berkwitz (1928) and Duffy (1930), rather than defining personality types in terms of amount of tension, added the criterion of variability. Whereas the phlegmatic individual was felt to exhibit a fairly constant tension level, the highly excitable person showed

a marked variation in tension level from one time to the next.

As supportive evidence for a personality theory based on differential activation of the skeletal muscular system, Duffy (1932a, 1957) and Arnold (1942, 1943) have attempted to show that individuals retain their respective ranks in amount of muscle tension displayed from one situation to the next. Although Duffy (1946) admits the existence of intraindividual differences due to variance in measurement techniques, nature of the task performed, and time at which data is gathered, she does not believe that such specific differences transcend an independent and general factor of muscle tension. Davis (1938) also believes that there are marked individual consistencies in muscular activity.

As compared to the more relaxed, individuals who are highly responsive in terms of their skeletal musculature also tend to be generally hyperactive. The tendency to respond to noise with a high increase in muscle tension was found by the present writer (Balshan, 1962) to be negatively related to Restraint and positively related to General Activity on the Guilford-Zimmerman Temperament Survey. The tension score was based on those muscles with the most significant loadings on a general muscle-tension factor. Muscle tension was defined by means of the electromyogram. There was no relationship with any of the other Guilford-Zimmerman temperaments.

Lundervold (1952) found that increased activity as shown by the electromyograph was correlated with a tendency to react to increases in noise, decreases in room temperature, and changes in other stimuli with anger and irritation. Similarly, Freeman and Katzoff (1932) revealed a positive correlation between grip-pressure scores and subjective ratings for annoyance and irritability.

The fact that individuals tend to respond in an idiosyncratic manner by means of their skeletal musculature was also indicated by Jacobson and Kraft (1942). On the basis of action potentials recorded during a period of reading, three types of responses became apparent. Some individuals showed an increase in muscle tension, others a decrease, while a third group responded with ups and downs throughout the entire period.

A rather unique personality breakdown in terms of response to stress was made by Kempe (1956). He found that those persons who tend to respond to stress by a rather general increase in muscle tension are able to deny emotion and to intellectualize quite easily. They remain aloof from others, maintaining very little regard for social convention. Contrasted with such types there are those individuals who, during stress, react by way of the autonomic nervous system. Autonomic responders are emotionally sensitive, prone to worry a great deal, and have fears of not being accepted by others.

There also appears to be some relationship between muscle tension and social adjustment (Lewis, 1937). Duffy (1932a) showed that the greater the degree of pressure applied to a dynamometer the lower were children rated by their nursery-school teachers on their degree of environmental adjustment. The more tense group was believed to be living under conditions of strain or effort.

Boys judged to be most tense were found by Wenger (1938) to be emotional, unstable, irritable, and aggressive. Using a larger sample of children Wenger (1943) obtained positive correlations between ratings of muscle tension and energeticness, frequency of emotion, carelessness, restlessness, distractibility, and impulsiveness. There

were negative correlations between the ratings and *fatigability*, *sensitiveness*, and emotional control.

Negative results were found by Martin (1958a, 1958b) who selected students with extreme scores in the following four groups: introverted neurotics, extraverted neurotics, introverted normals, and extraverted normals. Left eyelid activity and muscle-action potentials of the frontalis and the right forearm extensors were obtained from each of the four groups during rest and verbal activity. None of these measures was found to be related to the personality dimensions.

Just how much muscle tension varies from one situation to the next is still somewhat of an empirical question. The results of Voas (1952) indicate that the amplitude of action potentials from seven separate muscle groups vary not only for different individuals, but under different conditions in the same individual. While Malmö (1957) would probably concur with this view, he would also tend to state that there are some individuals who show higher levels of physiological reactivity than others to a larger number of different situations. What these situations have in common is that they all appear stressful to the individual. Only to the extent that such persons are able to avoid conditions with heightened arousal values, do they react similarly to other people.

#### RESPONSES TO EMOTIONAL STRESS

Almost all individuals react to emotional stress with some increase in action in the skeletal motor system. Emotional material produces changes in the frequency, amplitude, and duration of muscle-action potentials taken from the forehead (Newman, 1953), as well as increasing tension in the neck and in the forearm flexors and extensors (Klein, 1957). When severe emotional stimuli

were introduced there were increases in voluntary arm and hand movements which were closely related to the degree of induced excitement (Kellogg, 1932).

According to Duffy (1932b), the most extreme increases in muscle tension result from emotional stimuli, particularly when an individual is confronted with problems of adjustment. Not only does he apply increased effort at that time, but frequently his excitation level becomes too high and coordination of responses becomes more difficult.

Under increased strain Arnold (1942, 1943) found the more tense individual to break down relatively quickly due to an inability to utilize specific musculature in attending to tasks, thereby responding with a rather inefficient means of energy expenditure.

By means of pressure on tambours, Luria (1932) studied muscular reaction in response to distressing verbal stimuli. During a state of affect there was a disorganization of response, taking on an impulsive, excitatory character, accompanied by a tremor. In response to critical words not only did the individual apply more pressure to the tambour for his right hand, but this tended to overflow to the nonactive hand as well.

In a similar study, Runkel (1936) reported that reactions to words of a critical nature were in the form of wave-like tremors immediately preceding the motor response. The fact that there were no appreciable changes in the nonactive hand may have been due to the relatively low affect in the experimental situations.

Sudden increases in muscle tension are particularly apparent during the discussion of stressful material in psychiatric interviews. Not only are these increases revealed in the patient, but they can be found in the examiner as well. Whereas speech-muscle tension of the patient fell rapidly in response to

praise, it remained high when he was criticized. A similar response pattern was exhibited by the therapist (Davis & Malmo, 1951; Malmo, Boag, & Smith, 1957).

#### PSYCHIATRIC GROUPS

Under normal conditions, brief and temporary states of heightened tension commonly occur in response to problem-solving situations which require a certain degree of attentiveness. Under such conditions, however, the degree of tension necessary to keep the individual alert is usually not sufficient to call forth his total resources or to bring on pathological symptoms of overstimulation. On the other hand, there are certain individuals who are almost always extremely tense. They are no longer beneficially stimulated, but instead become hyperactive and restless. Accompanying the marked excitability of the muscular system are clinical manifestations of increased tendon reflexes, spasticity of smooth muscle, tremor, a slightly irregular respiration rate, and a somewhat increased pulse rate (Guyton, 1956; Jacobson, 1938).

Some (Freeman, 1939; Plutchik, 1954) claim that persistently high muscle-tension levels commonly exist in certain neurotic individuals who are in constant states of conflict. The conflict is thought to produce emotional reactions which are muscular preparations for action.

In an attempt to test the hypothesis that excessive stimulation of the skeletal motor system exists in psychological disorders, Henley (1935) compared normals with psychiatric groups, both psychotic and neurotic. With all groups combined, tension proved to be higher for men than women. For normal men it was higher than for patient groups of men, but was higher for female patients than normal women. With the introduc-

tion of a mental task there was an increase in the amount of tension in psychotic and neurotic women, while all of the other groups showed a decrease. When the psychiatric population was further subdivided, tension scores and their scatter about the mean were distributed in the following order from greatest to least: manic-manic, psychoneurotic, dementia praecox, and depressed-manic. Because of the questionable reliability of the measure used to define muscle tension (time taken by the forearm to cover a defined arc when moved by a stated weight and the outmoded diagnostic categories), the results are somewhat dubious.

More recently it has been shown that psychiatric patients exhibit increases in muscle tension under most conditions. Although stress usually brings about marked increments in tension, it is probable that to such individuals most situations are perceived as being stressful. During relatively simple voluntary motor tasks higher residual action potentials were recorded from forearm flexors and extensors of patients than from normal individuals (Davidowitz, Brown-Mayers, Kohn, Welch, & Hayes, 1955).

There were distinct differences between well-adjusted and emotionally unstable children in physiological measures taken during rest, attention, learning, frustration, recall, and sensory stimulation. While some differences were found to exist during both rest and attention, they were particularly prevalent during frustration. In the emotionally unstable group an exaggerated reaction was found in skin resistance, heart rate, blood pressure, respiratory irregularity, muscle tension (as measured by amount of tremor), and suppression of alpha rhythm. Whereas prepsychotic and neurotic individuals presented unstable patterns and reacted most violently to frustration, both

psychotic (schizophrenic) and well-adjusted persons showed patterns of greater stability (Jost, 1941; Sherman & Jost, 1942).

In a comparison of manic-depressives, schizophrenics, psychoneurotics, and normal adults during motor tasks, in contrast to the results of Sherman and Jost, it was the schizophrenics who showed the greatest increases in muscle tension (Wulfeck, 1941). The apparently contradictory results may be due to several factors. To begin with, Sherman and Jost only tested three schizophrenics. In addition, the only measure of muscle tension was the amount of tremor present. Moreover, the amount of physiological reactivity may be related to the chronicity of the psychosis.

As compared to 10 normal individuals a group of 21 schizophrenics were found by Whatmore and Ellis (1958) to have higher muscle-action-potential levels in the forehead, jaw, forearm, and leg. A similar result was reported by Petursson (1962) in a group of 10 patients, most of whom were schizophrenics.

During conditions of pain-stimulation, rapid discrimination, and mirror drawing, psychoneurotics, acute psychotics, and chronic schizophrenics, as compared to normals, responded by higher muscle tension in the neck and right arm. Such findings were identical for all three stress conditions. While muscle-tension level was as high in schizophrenics as in the other psychiatric patients, schizophrenics were less reactive if the pain stimulus was of brief duration (Malmo, Shagass, & Smith, 1951).

Very little work has been reported on depressed patients. Whatmore and Ellis (1959) took electromyographic recordings of the forehead, jaw, forearm, and leg in 19 female depressives, 16 of whom exhibited extreme motor retardation. When compared with a normal

sample of comparable age, the patient groups were characterized by higher muscle-tension levels. The magnitude of elevation, however, was greater in the more retarded group. When further investigations (Whatmore & Ellis, 1962) were carried out with five of the retarded patients, there was found to be a drop in muscle tension during psychiatric treatment. On the other hand, both before relapse and after treatment the muscle readings remained high. Such results would lead one to believe that high muscle tension is a general characteristic of the depressed person.

Impaired motor control and excessive muscular tension have been found to characterize psychoneurotic disorders, especially during stress. Malmö, Shagass, and Davis (1951) believe that during mirror drawing all individuals (normal and patient populations) respond with an equal "autonomic preparatory response." The difference arises, however, in that psychotics and psychoneurotics both have greater "skeletal muscle preparatory responses." This suggests a defect in the skeletal motor system, resulting in greater tensional output (Malmö, 1956; Malmö, Shagass, Belanger, & Smith, 1951).

Similarly, psychoneurotics were noted by Jurko, Jost, and Hill (1952) to over-respond to specific stimulation via the skeletal system and to have longer rates of recovery from stress. In addition to the increased blood pressure and lowered skin resistance with which normal individuals responded to stress, the psychoneurotics showed increases in body movement and respiration rate as well.

During agitated states neurotic patients evaluated by Jacobson (1952) exhibited increases in muscle tension in all four regions from which he recorded muscle action potentials: brachial biceps, quadriceps femoris, eyebrow, and jaw.

Martin (1956) compared muscle-

action potentials in psychotic, neurotic, and normal individuals from the forehead and right forearm during various stimulus conditions and during rest. While there were no differences in the resting tension levels of the neurotic and normal individuals, a difference was apparent between psychotic and normal persons, with psychotics being significantly more tense. In addition, "dysthymics" showed greater amounts of tension than a group of hysterics. Among the various stress situations, only when responding to questions about their symptoms did the neurotic group manifest higher levels of forearm tension. Differences between this study and those of Malmö may be due to differences in populations sampled.

In an investigation of muscle tension during handwriting in psychoneurotic and normal individuals it was found that the former group utilized both greater grip pressure on a stylus and more variability in grip pressure and point pressure. Time spent in writing was also greater among neurotics, and their relaxation time during writing was less. Such results support the hypothesis that greater muscular energy is expended by neurotics (Ruesch, Cobb, & Finesinger, 1951; Ruesch & Finesinger, 1943a, 1943b).

Due to the presence of muscular tension in so-called nervous individuals there is difficulty in relaxing. Lundervold (1951) has shown that individuals who are not so tense can relax in many positions and sit comfortably over long periods of time. More tense persons, on the other hand, are unable to relax in many positions and can only change with a marked increase in muscle strain.

Although the inability to relax is quite marked among psychiatric populations, Jacobson (1934) has found such a phenomenon also to be characteristic of a group of college students. After a

period of training the group of neurotic individuals could be taught to relax to a greater extent than the untrained college students (Jacobson, 1938).

### ANXIETY

Among the psychoneurotic disorders, one finds anxiety linked with muscle tension in the majority of clinical reports. Restlessness, strain, tension, and overexcitability are only a few of the terms commonly associated with anxiety states. The anxious person speaks of tension in his body and of mental tension. Yaskin (1937) defines anxiety as:

a form of affectivity recognized introspectively as an unpleasant affect without known cause, or with cause inadequately understood or mistaken, and manifested objectively by changes in neuromuscular, autonomic visceral and secretory functions (emotional expressions) [p. 9].

According to Jacobson (1942), tension in the skeletal muscles is a component, rather than a mere accompaniment of anxiety. The role of muscle tension in anxiety is suggested by a decrease in both tremor and anxiety following relaxants (Dickes, Flamen, Cottrera, & Tobin, 1955). In addition, Kueth and Eriksen (1957) claim that some of the effects of anxiety on performance may be due to muscle tension which is also present. Anxiety states are stated by Cameron (1944a) to originate in those persons who have a tendency to develop tension more readily than most individuals.

It is believed by Whitehorn (1944) that motor reactions are utilized by many neurotic patients to minimize their anxiety. Tics, embarrassed laughs, and obsessive rituals are forms of motor expression which serve as a release for the anxious person.

In a comparison of the responses to a pain stimulus in anxiety neurotics, schizophrenics, and controls, the degree

of physiological reactivity was related to the amount of anxiety present. Measures of motor function (finger-movement and neck-muscle-action potentials) were found to be related significantly to degree of disturbance and to amount of anxiety (Malmo, 1956; Malmo & Shagass, 1949a).

Eyeblink rate, an indirect index of tension, is felt by Meyer (1953) to be the best overall measure of generalized tension. Meyer, Bahrack, and Fitts (1953) and Lovaas (1960) found that individuals who score high on anxiety inventories also have high blink rates. Using muscle-action potentials in the arm as a measure of tension, Rossi (1959) also found a relationship between manifest anxiety scales and muscle tension.

Although Lewinsohn's (1956) group of anxious patients tended to show a greater tremor response to stress than a nonanxious control group, the difference was not significant. Both the anxious patients and a group of ulcer patients did, however, have significantly greater variance of finger-tremor scores. Salivary secretion and skin conductance did not appear to distinguish the anxious group from the controls.

The possibility exists that anxious and nonanxious individuals may be more easily differentiated by after responses to a stimulus than by immediate responses. After a reaction, it takes the anxious person some time before he returns to a basal level (Barlow, 1959).

Further support for such a hypothesis comes from a study by Malmo, Shagass, and J. F. Davis (1950). The initial tension level to a 100-cycle-per-second tone of 3 seconds duration, as measured by action potentials in the right forearm, was found to be equal in a control group and in one suffering from severe pathological anxiety. The control group returned in a short time to the previous level of tension; the anxious group

augmented the response and continued it during the entire period of auditory stimulation. As a result, the peak of reactivity for the patients occurred much later than for the controls.

When a burst of white noise was administered at 1-minute intervals, there was a somewhat larger initial response in anxiety patients than a group of controls (Davis, Malmo, & Shagass, 1954). A more significant difference, however, was the patients' higher forearm-extensor tension, .3-.4 seconds after stimulation, when the response of the control group had almost subsided.

There is some question as to whether anxious individuals are always muscularly tense, or if the elevation in tension only occurs in response to a stressful situation. In other words, if it were possible to get the anxious individual in a resting state, would he show any differences from a relatively nonanxious person? The present writer (Balshan, 1962) found that when women students were divided into high- and low-anxiety groups on the basis of the Taylor and Freeman Manifest Anxiety scales, there were no differences in their resting muscle-action potentials. This was true for all 16 of the skeletal muscles measured. In response to a white noise, however, the high-anxiety group reacted with significantly higher increases in muscle tension than the nonanxious one.

Malmo and Shagass (1949a, 1949b) feel that under conditions of stress there is a high probability that reaction tendencies of anxious patients will go directly into action. Doehring (1957) has also found anxiety to be highly correlated with certain tension reactions (eyeblick rate) in response to emotionally stressful words but not to relatively neutral words. A high score on a manifest anxiety scale is interpreted by Mednick (1957) as indicating that a person will show certain anxiety responses during traumatic situations.

Anxiety is not felt to be a chronic state which will manifest itself in any circumstance. The fact that anxious individuals are sometimes reported to maintain their high muscle-tension levels during rest may be due to the fact that the subjects were not truly relaxed but were reacting to the stress of the experiment itself.

#### SYMPTOM SPECIFICITY

While reaction to stress is frequently accompanied by increased activation of most of the skeletal musculature, there is increasing evidence that some individuals tend to exhibit high degrees of tension in specific muscle groups. In these individuals such a reaction is part of a life-long mechanism of responding to emotionally-toned events. It is frequently dependent not only on the present situation but on physiological constitution and past experience (Wolff, 1948).

After states of sustained muscular contraction in certain muscles, those muscles may become the source of pain and tenderness, outlasting the actual contraction. Wolff (1953) claims that headaches, backaches, periarticular pain, and aches of the extremities will develop in individuals who are characteristically tense and manifest inappropriate readiness for action.

Sustained muscular contraction secondary to noxious impulses, from disease of any structures in the head, or associated with emotional tension and anxiety, have been identified by Wolff as a potential source of headache and other head sensations. The intensity of the pain is positively correlated with muscle-action potentials from the head area. In cases of extreme apprehension the muscles contract as if preparing the individual for action; but the action may not occur. As a result, there is a persistent feeling of soreness radiating from the head and neck. Stimulation of the

pain nerve endings is felt to be the result, not only of sustained contraction, but of reduction of blood supply to the muscles caused by artery constriction and interference with blood flow (Simons, Day, Goodell, & Wolff, 1943; Wolf & Wolff, 1953; Wolff, 1948).

Those patients having head, neck, or arm symptoms were found by Sainsbury and Gibson (1954) to have higher action potentials in the relevant muscle than those without the symptom. It is also significant that the onset of headache was accompanied by corresponding increases in the activity of the frontalis muscle. Those more generalized bodily symptoms, such as tremor and startle, were accompanied by increases in muscle activity in well-separated areas.

In an investigation of 400 patients, Friedman and Von Storch (1953) found what they called tension headache to be predominantly an emotionally produced physiological reaction to anxiety-provoking situations. Although relief was obtained by placebo in 50% of the cases, long-lasting effects were achieved only through psychotherapy.

Other individuals tend to develop backaches due to sustained contractions in the lumbar sacro spinalis, trapezius, and hamstring muscles. Situations which either threaten security or impend conflict, guilt, or anxiety have been found to produce an increase in electromyographic activity in patients with a history of backache (Holmes & Wolff, 1950).

Similar observations have been made by Malmö, Shagass, and Davis (1950b) on a group of headache-prone patients. Whenever traumatic material was discussed in interview sessions, such a patient reported the presence of a headache and showed a rise in muscle-action potentials recorded from the forehead.

In response to a series of painful stimulations elicited by a Hardy-Wolff thermal stimulator, patients with head

and neck pain again exhibited increases in muscle-action potentials from these areas. In contrast, those patients with clinical histories of cardiovascular complaints showed increases in heart rate, mean heart rate variability, and median respiratory variability, rather than displaying any increases in muscular activity (Malmö & Shagass, 1949b; Malmö, Shagass, & Davis, 1950a).

In order to explain specific muscular effects Malmö, Shagass, and Davis (1950b) speak of "symptom specificity." They believe that a given patient will respond maximally with the same physiological response, almost regardless of the stimulus. Accordingly, a particular physiological mechanism of somatic complaint is specifically susceptible to activation by many stressful situations.

Evidence also has been gathered to show that specific bodily changes occur, not only in response to general emotional disturbances, but to particular emotional conflicts as well. Certain psychodynamic themes occurring during psychiatric interviews appear to be associated with increases in muscle tension in rather circumscribed areas. Hostile material brings about an increase in forearm tension, while sexual themes produce a rise in leg tension (Shagass & Malmö, 1954). It may be seen that certain patterns during emotional stress are exhibited by various individuals. Why some develop backaches or headaches, others more diffuse patterns, and others still different physiological patterns is not known at present. The particular mode of response, however, usually appears to be part of a life-long means of reaction. One might search for the precipitating factor in the weakness of an organ, which when activated becomes especially susceptible to stimulation during stress. At the same time, environmental factors cannot be completely ignored; for certain behavior would probably not continue if it were in no way reinforced.

It is particularly interesting that many persons with the same manifestations during stress have common experiences (Wolff, 1953). It seems possible that bodily reactions conditioned by earlier exposure occur in similar situations encountered at a later date.

Concerning symptom specificity, a word of warning is introduced. Because an individual's tension may become focused in a particular area, Cameron (1944b) believes that a general increase in tension may be overlooked. A certain area may have acquired special meaning for an individual and is activated during anxiety-provoking situations. While for some muscle tension is usually confined to the problem area, during extremely intense anxiety states it may extend to other body parts and even to other systems. In addition, neuromuscular contractions in particular body parts may be so intense that general feelings of tension may not be as obviously noticed.

#### SUMMARY AND CONCLUSIONS

In scanning the literature it becomes obvious that the definitions of muscle tension are as numerous as are the methods used to measure it. Some of the most indirect methods include ratings of tension, magnitude of knee jerk, eye-blink rate, motor restlessness, and tremor. Wendt (1938) notes that in most cases the correlation of these measures with more direct indices of muscle tension has not been investigated. When investigations have been made the correlations are often very low, making it very likely that the various methods convey different results. Even where the electromyogram is utilized to assess tension levels, different muscles have been measured. In many cases recordings are made of only one or two muscle groups, which have not always been representative of generalized muscle tension.

In spite of the frequent lack of comparability of studies, it is still possible to draw some generalizations and to perceive the emergence of certain trends. To begin with, distinct differences in tension level exist between individuals and tend to be maintained under many situations. There are also suggestions that muscle tension relates to a variety of traits, although the relationship is not always clear-cut. The common factor amongst these traits appears to be a general tendency to be extremely responsive to most stimuli. Accompanying the high muscle tension would be the tendency to act out and express emotion freely (i.e., to be more irritable, impulsive, energetic, active, excitable).

When the situation becomes a particularly stressful one, muscle-tension levels become elevated in almost all people, although certain individuals would still be more responsive than others. Among those individuals characterized by high frequency of elevated motor responses are psychiatric populations, particularly those with symptoms of anxiety.

There is very little agreement in the literature as to which of the psychological disorders is accompanied by the greatest amount of muscle disturbance. This may be accounted for partially by the noncomparability of the experimental situations. In addition, there is tremendous variation in the methodology of assessing muscle tension, as well as in the decision regarding which muscle responses to measure. Another problem arises because of the unreliability of the diagnostic categories. Finally, the chronicity of the psychosis or neurosis may have an important bearing on the magnitude of muscle tension.

In addition to those individuals who respond with increases in tension throughout their body, some people exhibit high degrees of tension in a specific muscle group. The ideas of both specific and general tension are not nec-

essarily exclusive; for it is possible for the same person to be tense throughout his body and yet respond to most situations with the greatest amount of tension being focused in one particular muscle. Sustained muscle contraction in turn may lead to pain and symptoms in the corresponding muscle; so that the individual reports headaches or backaches. Exactly why one muscle becomes so significant is due not only to the present situation, but to the physiological constitution of the person and to his past experiences as well.

## REFERENCES

- ALLPORT, G. W., & VERNON, P. E. *Studies in expressive movement*. New York: Macmillan, 1933.
- ARNOLD, M. R. A study of tension in relation to breakdown. *J. gen. Psychol.*, 1942, 26, 315-346.
- ARNOLD, M. B. A study of tension in relation to breakdown. *Pap. Amer. Congr. Gen. Semant.*, 1943, 2, 209-220.
- BALSILIAN, I. D. Muscle tension and personality in women. *Arch. gen. Psychiat.*, 1962, 7, 436-448.
- BARLOW, W. Anxiety and muscle-tension pain. *Brit. J. clin. Pract.*, 1959, 339-350.
- BARTOSHUK, A. Electromyographic gradients as indicants of motivation. *Canad. J. Psychol.*, 1955, 9, 215-230.
- BILLS, A. G. Facilitation and inhibition in mental work. *Psychol. Bull.*, 1937, 34, 286-309.
- BROMAN, T. Electromyo-mechanographic registrations of passive movements in normal and pathological subjects. *Acta psychiat. Copenhagen*, 1949, Suppl. No. 53.
- CAMERON, D. E. Autonomy in anxiety. *Psychiat. Quart.*, 1944, 18, 53-60. (a)
- CAMERON, D. E. Observations on the patterns of anxiety. *Amer. J. Psychiat.*, 1944, 101, 36-41. (b)
- COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, 39, 347-367.
- DAVENPORT, C. B. *Feebly inhibited: Nomadism or the wandering impulse with special reference to heredity; inheritance of temperament*. (Publ. No. 236) Pittsburgh, Pa.: Carnegie Institution, 1915.
- DAVIDOWITZ, J., BROWN-MEYERS, A. N., KOHN, R., WELCH, L. L., & HAYES, R. An electromyographic study of muscular tension. *J. Psychol.*, 1955, 40, 85-94.
- DAVIS, F. H., & MALMO, R. B. Electromyographic recording during interview. *Amer. J. Psychiat.*, 1951, 107, 908-916.
- DAVIS, J. F., MALMO, R. B., & SHAGASS, G. Electromyographic reaction to strong auditory stimulation in psychiatric patients. *Canad. J. Psychol.*, 1954, 8, 177-186.
- DAVIS, R. C. The relation of muscle action potentials to difficulty and frustration. *J. exp. Psychol.*, 1938, 23, 141-158.
- DAVIS, R. C. Methods of measuring muscular tension. *Psychol. Bull.*, 1942, 39, 329-346.
- DEUTSCH, F. Analysis of postural behavior. *Psychoanal. Quart.*, 1947, 16, 195-211.
- DICKES, R., FLAMEN, G. H., COTTERERA, J., & TOBIN, M. The effect of mephenesin in muscle tension: An experimental study. *AMA Arch. Neurol. Psychiat.*, 1955, 74, 590-597.
- DOEHRRING, D. G. The relation between manifest anxiety and rate of eyeblink in a stress situation. *USN Sch. Aviat. Med. res. Rep.*, 1957, No. 6.
- DOWNNEY, J. *The will-temperament and its testing*. New York: World Book, 1923.
- DUFFY, E. Tension and emotional factors in reaction. *Genet. Psychol. Monogr.*, 1930, 7, 1-79.
- DUFFY, E. The measurement of muscular tension as a technique for the study of emotional tendencies. *Amer. J. Psychol.*, 1932, 44, 146-162. (a)
- DUFFY, E. The relation between muscular tension and quality of performance. *Amer. J. Psychol.*, 1932, 44, 535-546. (b)
- DUFFY, E. Level of muscular tension as an aspect of personality. *J. gen. Psychol.*, 1946, 35, 161-171.
- DUFFY, E. The psychological significance of the concept of arousal or activation. *Psychol. Rev.*, 1957, 64, 265-275.
- DUNBAR, F. *Emotions and bodily changes*. New York: Columbia Univ. Press, 1946.
- FREEMAN, G. L. Postural tensions and the conflict situation. *Psychol. Rev.*, 1939, 46, 226-240.
- FREEMAN, G. L. *Physiological psychology*. New York: Van Nostrand, 1948.
- FREEMAN, G. L., & KATZOFF, E. T. Muscular tension and irritability. *Amer. J. Psychol.*, 1932, 44, 789-792.
- FRIEDMAN, A. P., & VON STORCH, T. Tension headache. *J. Amer. Med. Ass.*, 1953, 151, 174-177.
- GUYTON, A. C. *Textbook of medical physiology*. Philadelphia, Pa.: Saunders, 1956.
- HAMMETT, F. S. Observations on the relation between emotional and metabolic stability. *Amer. J. Physiol.*, 1920, 53, 307-311.
- HENLEY, E. H. Factors related to muscular

- tension. *Arch. Psychol.*, N. Y., 1935, 27, 1-44.
- HOLMES, T. H., & WOLFF, H. G. Life situations, emotions, and backaches. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1950, 29, 750-772.
- JACOBSON, E. The electrophysiology of mental activities. *Amer. J. Psychol.*, 1932, 44, 677-694.
- JACOBSON, E. Electrical measurements concerning muscular contraction (tonus) and the cultivation of relaxation in man: Studies on arm flexors. *Amer. J. Physiol.*, 1934, 107, 230-248.
- JACOBSON, E. *Progressive relaxation*. Chicago: Univer. Chicago Press, 1938.
- JACOBSON, E. The physiological conception and treatment of certain anxiety states. *Psychol. Bull.*, 1942, 39, 599.
- JACOBSON, E. Specialized electromyography in supplement to clinical observations during hyperkinetic states in man ("functional nervous conditions"). *Federal Proc.*, 1952, 11, 77.
- JACOBSON, E., & KRAFT, F. L. Contraction potentials (right quadriceps femoris) in man during reading. *Amer. J. Physiol.*, 1942, 137, 1-5.
- JAMES, W. *The principles of psychology*. Vol. 2. New York: Holt, 1890.
- JOHNSON, B. Changes in muscular tension in coordinated hand movements. *J. exp. Psychol.*, 1928, 11, 329-341.
- JOST, H. Some physiological changes during frustration. *Child Develpm.*, 1941, 12, 9-15.
- JURKO, M., JOST, H., & HILL, T. S. Pathology of the energy system: An experimental clinical study of physiological adaptive capacities in a nonpatient, a psychoneurotic, and an early schizophrenic group. *J. Psychol.*, 1952, 33, 183-198.
- KELLOGG, W. N. The effect of emotional excitement on muscular steadiness. *J. exp. Psychol.*, 1932, 15, 142-166.
- KEMPE, J. E. An experimental investigation of the relationship between certain personality characteristics and physiological responses to stress in a normal population. Unpublished doctoral dissertation, Michigan State University, 1956.
- KENNEDY, J. L., & TRAVIS, R. C. Prediction and control of alertness: II. Continuous tracking. *J. comp. physiol. Psychol.*, 1948, 41, 203-210.
- KLEIN, A. Variations in muscle tension during mental activities involving "emotional" material. Unpublished doctoral dissertation, New York University, 1957.
- KUETHE, J. L., & ERIKSEN, C. W. Personality, anxiety, and muscle tension as determinants in response stereotypy. *J. abnorm. soc. Psychol.*, 1957, 54, 400-404.
- LEWINSOHN, P. M. Physiological reactivity to stress. *J. comp. physiol. Psychol.*, 1956, 49, 271-277.
- LEWIS, F. H. Muscular tension and social adjustment in pre-adolescent children. *Psychol. Bull.*, 1937, 34, 781-782.
- LOVAAS, O. I. The relationship of induced muscular tension, tension level, and manifest anxiety in learning. *J. exp. Psychol.*, 1960, 59, 145-152.
- LUNDERVOLD, A. Electromyographic investigations during sedentary work, especially type-writing. *Brit. J. phys. Med.*, 1951, 14, 32-36.
- LUNDERVOLD, A. An electromyographic investigation of tense and relaxed subjects. *J. nerv. ment. Dis.*, 1952, 115, 512-525.
- LURIA, A. R. *The nature of human conflicts*. New York: Liverwright, 1932.
- MCKINLEY, J. C., & BERKOWITZ, N. J. Quantitative studies in human muscle tonus. *Arch. Neurol. Psychiat.*, 1928, 19, 1036-1056.
- MALMO, R. B. Symptom mechanisms in psychiatric patients. *Trans. N. Y. Acad. Sci.*, 1956, 18, 545-549.
- MALMO, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, 64, 276-287.
- MALMO, R. B., BOAG, T. J., & SMITH, A. A. Physiological study of personal interaction. *Psychosom. Med.*, 1957, 19, 105-119.
- MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, 11, 9-24. (a)
- MALMO, R. B., & SHAGASS, C. Physiologic study of symptom mechanisms in psychiatric patients under stress. *Psychosom. Med.*, 1949, 11, 25-29. (b)
- MALMO, R. B., SHAGASS, C., BELANGER, D. J., & SMITH, A. A. Motor control in psychiatric patients under experimental stress. *J. abnorm. soc. Psychol.*, 1951, 46, 539-547.
- MALMO, R. B., SHAGASS, C., & DAVIS, F. H. A physiological study of somatic symptom mechanisms in psychiatric patients. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1950, 29, 23-261. (a)
- MALMO, R. B., SHAGASS, C., & DAVIS, F. H. Symptom specificity and bodily reactions during psychiatric interview. *Psychosom. Med.*, 1950, 12, 362-376. (b)
- MALMO, R. B., SHAGASS, C., & DAVIS, J. F. A method for the investigation of somatic response mechanisms in psychoneurosis. *Science*, 1950, 112, 325-328.
- MALMO, R. B., SHAGASS, C., & DAVIS, J. F. Electromyographic studies of muscular ten-

- sion in psychiatric patients under stress. *J. clin. exp. Psychopathol.*, 1951, 12, 45-66.
- MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *J. Pers.*, 1951, 19, 359-375.
- MARTIN, I. Levels of muscle activity in psychiatric patients. *Acta psychol., Amsterdam*, 1956, 12, 326-341.
- MARTIN, I. Blink rate and muscle tension. *J. ment. Sci.*, 1958, 104, 123-132. (a)
- MARTIN, I. Personality and muscle activity. *Canad. J. Psychol.*, 1958, 12, 23-30. (b)
- MEDNICK, S. A. Generalization as a function of manifest anxiety and adaptation to psychological experiments. *J. consult. Psychol.*, 1957, 21, 491-494.
- MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, 50, 204-220.
- MEYER, D. R., BAHRICK, H. P., & FITTS, P. M. Incentive, anxiety, and the human blink rate. *J. exp. Psychol.*, 1953, 45, 183-187.
- NEWMAN, P. P. EMG studies of emotional states in normal subjects. *J. Neurol. Neurosurg. Psychiat.*, 1953, 16, 200-208.
- PETURSSON, E. Electromyographic studies of muscular tension in psychiatric patients. *Comprehen. Psychiat.*, 1962, 3, 29-36.
- PLUTCHIK, R. The role of muscular tension in maladjustment. *J. gen. Psychol.*, 1954, 50, 45-62.
- ROSSI, A. M. An evaluation of the manifest anxiety scale by the use of electromyography. *J. exp. Psychol.*, 1959, 58, 64-69.
- RUESCH, J., COBB, S., & FINESINGER, J. E. Studies on muscular tension in the neuroses. *Trans. Amer. Neurol. Ass.*, 1951, 67, 186-189.
- RUESCH, J., & FINESINGER, J. E. Muscular tension in psychiatric patients. *Arch. Neurol. Psychiat.*, 1943, 50, 439-449. (a)
- RUESCH, J., & FINESINGER, J. E. The relation between electromyographic measurements and subjective reports of muscular relaxation. *Psychosom. Med.*, 1943, 5, 132-138. (b)
- RUNKEL, J. E. Luria's motor method and word association in the study of deception. *J. gen. Psychol.*, 1936, 15, 23-27.
- SAINSBURY, P., & GIBSON, J. G. Symptoms of anxiety and tension and the accompanying physiological changes in the muscular system. *J. Neurol. Neurosurg. Psychiat.*, 1954, 17, 216-224.
- SHAGASS, C., & MALMO, R. B. Psychodynamic themes and localized muscular tension during psychotherapy. *Psychosom. Med.*, 1954, 16, 295-314.
- SHERMAN, M., & JOST, H. Frustration reac-
- tions of normal and neurotic persons. *J. Psychol.*, 1942, 13, 3-19.
- SIMONS, D. J., DAY, E., GOODELL, H., & WOLFF, H. G. Experimental studies on headache: Muscles of the scalp and neck as sources of pain. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1943, 23, 228-244.
- STROUD, J. B. The role of muscular tension in stylus maze learning. *J. exp. Psychol.*, 1931, 14, 606-631.
- TEICHNER, W. H. Effects of foreperiod, induced muscular tension, and stimulus regularity on simple reaction time. *J. exp. Psychol.*, 1957, 53, 277-284.
- VOAS, R. B. *Generalization and consistency of muscle tension level*. Unpublished doctoral dissertation, University of California, Los Angeles, 1952.
- WALLERSTEIN, H. An electromyographic study of attentive listening. *Canad. J. Psychol.*, 1954, 8, 228-238.
- WENDT, G. R. Methods of recording action. *Arch. Psychol., N. Y.*, 1938, 32, 1-83.
- WENGER, M. A. Some relationships between muscular processes and personality and their factorial analysis. *Child Developm.*, 1938, 9, 261-276.
- WENGER, M. A. An attempt to appraise individual differences in level of muscular tension. *J. exp. Psychol.*, 1943, 32, 213-225.
- WHATMORE, G. B., & ELLIS, R. M. Some motor aspects of schizophrenia. *Amer. J. Psychiat.*, 1958, 114, 882-889.
- WHATMORE, G. B., & ELLIS, R. M. Some neurophysiologic aspects of depressed states. *Arch. gen. Psychiat.*, 1959, 1, 70-80.
- WHATMORE, G. B., & ELLIS, R. M. Further neurophysiologic aspects of depressed states. *Arch. gen. Psychiat.*, 1962, 6, 243-253.
- WHITEHORN, J. C. Guide to interviewing and clinical personality study. *Arch. Neurol. Psychiat.*, 1944, 52, 196-216.
- WOLF, S., & WOLFF, H. G. *Headaches: Their nature and treatment*. Boston, Mass.: Little, Brown, 1953.
- WOLFF, H. G. *Headache and other head pain*. New York: Oxford Univer. Press, 1948.
- WOLFF, H. G. *Stress and disease*. Springfield, Ill.: Charles C Thomas, 1953.
- WOLFF, W. Involuntary self-expression in gait and other movements. *Charact. Pers.*, 1933, 2, 328-344.
- WULFECK, W. H. Motor function in the mentally disordered. Part 2. *Psychol. Rec.*, 1941, 4, 326-348.
- YASKIN, J. C. The psychobiology of anxiety. *Psychoanal. Rev.*, 1937, 24(Suppl.), 1-93.

(Received April 6, 1963)

# THE PSYCHOTHERAPEUTIC RELATIONSHIP

G. GAIL GARDNER<sup>1</sup>

*Teachers College, Columbia University*

This paper attempts an exhaustive review of the literature on the psychotherapeutic relationship. The nature of the ideal psychotherapeutic relationship is defined and approximations of this ideal are shown to correlate positively with various measures of patient progress in the therapeutic enterprise. Other correlates of good therapeutic relationships are then explored: patient variables, therapist variables, therapist-patient similarity, and technique variables. In all these areas, significant correlations are found, though not as often as one might expect. The research on therapist variables is especially disappointing, while patient variables seem to be quite good predictors of the quality of the ensuing patient-therapist relationship. Methodological issues are considered throughout the paper.

Psychotherapy has been variously defined. In this paper, it will refer to

a warm, permissive, safe, understanding, but limited social relationship within which therapist and patient discuss the affective behavior of the latter, including his ways of dealing with his emotionally toned needs and the situations that give rise to them [Shoben, 1953, p. 127].

Some writers on the subject of psychotherapy have focused on the "within which" aspects, citing numerous techniques as being more or less beneficial. Others have focused on the fact that psychotherapy is a relationship, and they have asserted that factors directly associated with this phenomenon contribute significantly to success or failure. This paper addresses itself to the validity of the latter assertion. The literature cited covers the period 1946-62.

## CONCEPTIONS OF THE IDEAL THERAPEUTIC RELATIONSHIP

Since a relationship may be characterized in a myriad of ways, anyone who asserts that the relationship itself is important in psychotherapy must state which sorts of relationship are desirable and which not. As Shoben (1949)

pointed out, there is considerable agreement on this issue. The characteristics most frequently cited as desirable are the therapist's warmth, acceptance, permissiveness, respect for the patient, understanding, interest in the patient, and liking for the patient. Rogers (1957, 1959) made the additional stipulation that, in successful therapy, the patient must be able to perceive these therapist qualities, and he also asserted (Rogers, 1954) that the patient must like and respect the therapist.

A number of scales which have been constructed (Anderson & Anderson, 1954; Apfelbaum, 1958; Chase, 1946; Fiedler, 1950b; McClelland & Sinaiko, 1950; Sundland & Barker, 1962) for the purpose of defining the ideal relationship indicate that, while the factors cited above are the modal conception, agreement is far from perfect. Thus the issue is raised as to what factors are related to agreement with the modal conception.

Chase (1946) derived his scale from statements about counseling procedure which were endorsed by a majority of "expert" counselors. Counseling students' attitudes generally did not agree with those of the experts and did not correlate either with grades in the counseling course or with Army General Classification Test scores. Chase concluded

<sup>1</sup> I wish to express gratitude to Allen E. Bergin, Rosalea A. Schonbar, and Laurance F. Shaffer for their helpful comments and criticisms during the preparation of this paper.

that acquisition of effective counseling attitudes was not related to scholastic achievement and probably was a function of actual experience in the counseling situation.

McClelland and Sinaiko (1950) refined the Chase scale and administered it to undergraduate and graduate psychology students. Comparisons within each of the two levels confirmed Chase's finding regarding scholastic achievement. Between-group comparisons, however, showed higher scores for the graduate students, and the authors concluded that training helped the students to acquire their instructors' attitudes. Of course this discrimination may have been more a function of self-selection of graduate students than of the training itself. Arbuckle (1956) found high agreement among graduate students' descriptions of the ideal counselor. In both studies, however, it is a moot question whether the graduate students actually believed in the attitudes they professed or merely were more adept at judging what answer was expected.

Fiedler (1950b), using a *Q* sort technique, found for seven therapists representing three schools that there were no significant differences in conception of the ideal therapeutic relationship between therapists of different schools, but that experienced and inexperienced therapists of the same school did differ significantly from each other. He argued that ability to describe the ideal therapeutic relationship was a function of experience rather than of theoretical allegiance. Though Fiedler's evidence has often been cited as conclusive, it has not gone unchallenged. Apfelbaum (1958) criticized Fiedler's *Q* deck on the grounds that the items reflected extreme positions and thus tended to compel agreement among sorters, resulting in spuriously high intersorter correlations.

Behar and Altrocci (1961), using a

scale constructed by Apfelbaum, asked nursing students to describe the ideal psychiatric nurse. Participation in psychiatric nursing courses seemed to produce high agreement, whereas actual experience with psychiatric patients did not. The authors concluded that they had refuted Fiedler's (1950b) hypothesis concerning experience, and that training instead was the critical variable. However, they failed to note that the high-agreement groups had had considerable experience with nonpsychiatric patients. Since it is quite doubtful that student nurses are taught to hold significantly different attitudes toward psychiatric patients as opposed to patients in general, the authors' conclusions are questionable. The training and experience variables were not properly controlled in either this study or Fiedler's (1950b); thus the issue remains unresolved as to which contributes more to agreement on good therapeutic attitudes.

Sundland and Barker (1962) constructed a scale which did differentiate between therapeutic schools in terms of therapeutic attitudes. The criterion for item selection was that the items described points of controversy between schools; thus all that can be concluded is that the authors successfully validated their scale. In comparing their results to Fiedler's the authors recognized that different attitudes were measured, and they properly raised the issue of the extent to which differential responses to a scale may be a function of its content rather than of the persons responding. A related issue concerns the appropriate degree of specificity of items in an attitude scale. While everyone may agree that a therapist should be warm and accepting, there may be considerable disagreement as to what these terms actually imply. Sundland and Barker (1962) found, contrary to Fiedler's (1950b) results, that experienced and inexperienced

therapists did not differ in their responses. But it must be noted that if a therapist is trained in a given school of thought, he does not need much experience to learn his school's position on the more controversial issues. Thus Fiedler's findings regarding experience and schools have still not been conclusively refuted.

Three studies (Anderson & Anderson, 1954; Fiedler, 1950b; Thomas, Polansky, & Kounin, 1955) noted that persons with no professional experience or training could describe the ideal therapeutic relationship about as well as therapists. Fiedler hypothesized that the therapeutic relationship may be only a variation of good interpersonal relationships in general.

Soper and Combs (1962), using a modification of Fiedler's (1950b) Q deck, found that teachers described the ideal teacher in much the same way that expert therapists described the ideal therapist. These data cannot be said to confirm Fiedler's hypothesis that the therapeutic relationship is only a paradigm of good human relationships generally, but they do support the notion of commonality among helping relationships, at least as described in Fiedler's terms by experts in the respective fields.

About the most that can be concluded from these various scales and samples is that the ideal therapeutic relationship may be as uniformly described as the scale provided permits. It is probably a variant of helping relationships in general and can probably be reasonably well described by anyone who considers it important. This latter hypothesis might explain the fact that Anderson and Anderson's (1954) patients were successful in the task while Behar and Altrocci's (1961) nonnursing students failed. Personal commitment to the significance of helping relationships may be a more pertinent variable than experience per se. Professional training does not seem to be especially relevant.

## RELATIONSHIP FACTORS AND THERAPEUTIC CHANGE

We now consider studies in which aspects of the patient-therapist relationship are related to therapeutic change. The studies will be discussed in blocks according to the way in which the relationship variable is measured.

### *Therapist Ratings*

It is difficult to obtain independent measures of the two variables under consideration. When one judges a helping relationship to be good in that it satisfies the conditions outlined above, one tends to experience a feeling of satisfaction which in turn acts as a set for perceiving progress toward therapeutic goals. Similarly, the converse is true. Even when one variable is measured alone, it is difficult not to get cues regarding the other. Given this inherent contamination, it only adds fuel to the fire to propose a design in which the same people rate both variables. The following two studies suffer from this error, and their results should be interpreted with caution.

Gorlow, Hoch, and Telschow (1952) studied therapy groups and found that the therapists had greatest liking for the most profited group. Since no statistical data were provided, it cannot be ascertained how strong the relationship was between liking and progress. Seeman (1954) found that therapist ratings of liking for their patients correlated .65 with therapist ratings of therapeutic success ( $N=23$ ). Retest reliability of the ratings was satisfactory. Since both ratings were made at the end of therapy, contamination is very likely. Therapist ratings of the degree to which the patient used his relationship to the therapist as a focus for therapy were not related to success. This finding casts doubt on the validity of notions to the contrary held by many therapists.

Snyder (1961) had significantly more

positive affect toward his more improved patients. But the facts that his relationship and progress measures overlapped and that the patients were also his students suggest caution in interpreting these data and others from the same research to be reported below.

Two studies (Coons, 1957; Lesser, 1961) utilized projective and other personality tests to measure therapeutic change and thereby minimized the possibility of contamination. Coons (1957) found that a therapy group which focused on interaction showed significantly more improvement than either a group that stressed insight into psychological problems or a no-therapy control group. The latter two groups did not differ from each other. Coons concluded that interaction is the essential condition for therapeutic change. These findings contradict those of Seeman (1954). The difference may be due to the fact that Coons' patients were hospitalized schizophrenics and Seeman's were neurotics. Since withdrawal from interpersonal interaction is a major symptom of schizophrenia, whereas it may or may not be a central issue in neurosis, it seems entirely reasonable that measures of improvement would be more closely related to interaction in the former case than in the latter. The fact that Seeman's patients were in individual therapy and Coons' were in group therapy may also contribute to the difference in results.

Lesser (1961) found no relationship between therapists' ratings of their own empathic understanding and differences in patients' self-ideal discrepancies (Butler & Haigh, 1954) before and after therapy. It should be noted, however, that the number of hours of therapy varied from 3 to 12. For most patients in treatment for this short a period, even the most expert therapists cannot hope to effect significant changes in the more permanent personality constructs measured by the Butler and Haigh *Q* sort.

In all these studies, of course, the problems inherent in self-reports are relevant.

### *Patient Ratings*

Studies in which patients made both judgments (Feifel & Eells, 1962; Grigg & Goodstein, 1957; Lipkin, 1948) also suffer from interdependent measures and must therefore be interpreted with caution. A patient who feels warmly toward his therapist has more reason to judge himself improved; conversely, one who feels himself improved is more likely to have positive feelings toward his therapist, especially in retrospect.

Lipkin (1948) found that patients who mentioned the counselor's permissive attitude in posttherapy evaluations usually felt that it helped them to achieve insight into their problems. Though the length of treatment was no greater than that employed by Lesser (1961), the improvement criteria were much more situational; this difference probably explains the discrepant results. Grigg and Goodstein (1957), using a follow-up questionnaire, found that clients who felt a close relationship with their counselor reported significantly more favorable outcomes than did clients who felt a more distant relationship. Feifel and Eells (1962) obtained similar results.

Van der Veen (1961) utilized ratings by outside judges as the measure of improvement. He found significant positive correlations between patient ratings of the degree to which therapists provided a positive relationship and judges' process ratings.

A few studies have relied on test scores as the measure of progress in therapy. Lipkin (1954) measured therapy change by global judgments of Thematic Apperception Test protocols. He found that successful patients focused less than unsuccessful patients on their willingness to confide in the therapist, and when they

did so focus, it was in a more positive way. The unsuccessful patients focused on mistrust. Lesser (1961) found no relationship between client ratings of therapist empathy and changes in self-ideal discrepancies. These negative results have been explained in the section on therapist ratings. Snyder (1961) reported a significant positive relationship between patients' attitudes toward the therapist and their classification as better or worse patients. Since, however, the measure of the two variables overlapped, the results are contaminated.

### *Ratings by Outside Judges*

In studies in which judges rated both therapeutic change and the quality of the therapeutic relationship, it is not always clear whether or not the same judges made both ratings. Even if they did, these studies have merit over those described above in that the patient-therapist relationship is now rated by people who are not personally involved in the outcome of therapy.

Holt and Luborsky (1952) found that resident psychiatrists rated by their supervisors as successful were also rated as having better relationships with patients than were residents rated as unsuccessful. Unfortunately, the ratings were contaminated by the fact that the supervisors themselves had better relationships with the residents they judged to be successful. Luborsky (1952) noted that the successful residents were more flexible with respect to the range of behavior which they felt appropriate for their patients. Apparently the less successful residents tended to try to cast their patients into a preconceived mold. The finding that the successful residents had better relationships with ward personnel and fellow residents as well as with supervisors and patients lends some support to Fiedler's (1950b) hypothesis that good therapy relationships are not

unique entities but rather instances of good relationships in general. Parloff's (1956) findings further support this notion.

Knupfer, Jackson, and Krieger (1959) noted the authoritarian nature of the relationship between the supervisor and the beginning therapist, and argued that supervisors' ratings may be heavily weighted with evaluations of the therapists' attitudes toward authority. Nevertheless, such ratings are probably more valid than the therapists' own ratings.

Van der Veen (1961) reported greater process-movement scores for clients whose therapists were judged to create better relationships. Truax (1961a) reported similar results and properly pointed out that causality had not been demonstrated. One study (Parloff, 1961), in which it was made clear that the judges of the relationship were different from the judges of therapeutic change, reported significant positive correlations between the two variables for patients in group therapy.

Two studies (Hiler, 1958; Parloff, 1961) relied on patient reports and behavior as the measure of change. Parloff (1961) noted a positive correlation between judged quality of the therapeutic relationship and patient reports of relief from symptomatic discomfort. In Hiler's (1958) study, therapists rated by staff psychologists as warm were better able to keep unproductive patients in treatment.

Truax (1961b, 1962) measured change by various test scores, including the Minnesota Multiphasic Personality Inventory (MMPI), and reported that, for two samples of schizophrenic patients, judged therapist empathy was positively related to improvement.

Three studies measured therapy change with a combination of ratings and test scores. Aronson (1953) reported no differences in improvement for

clients of four therapists who were judged by their peers to have significantly different degrees of ability for warm interpersonal relationships. The fact that both therapists and judges were graduate students and that the judgments were not limited to therapist-patient relationships may have contributed to the null results. Truax (1961b), in a similarly designed study, obtained positive results for both neurotic and schizophrenic samples. His research differed from Aronson's primarily in that his relationship judgments were based on actual therapy sessions. Stoler's (1961) judges did not concern themselves with the therapists, but rather rated the likability of the patients on the basis of tape-recorded segments of therapy. Patients identified as more likable were more often identified as improved.

In evaluating all these studies, one must consider that, when there is little agreement as to the nature of improvement and no conclusive evidence that any form of change occurs as a result of psychotherapy, it seems premature blithely to examine the correlates of therapeutic change. Yet there seems to be no other appropriate course of action.

The evidence that the quality of the therapeutic relationship is a correlate of therapeutic change lies not in the conclusive results of any one study but rather in the repeated findings of a series of studies, most of which contain one or more serious defects. Methodology varies greatly, and absence of precise definitions often makes it difficult to discern whether the "good relationship" of one study contains the same elements as that of another study or different. In a sense, the diversity of procedure strengthens the force of the conclusion. Null results do not cluster in any one methodological cell, and, in all types of design, positive results occur far more frequently. Further support derives from

the high agreement among different types of raters regarding the nature of the relationship between a given therapist and patient. Miller (1949) reported nonsignificant differences between the judgments of four raters with widely varying amounts of training. Parloff (1961) and Snyder (1961) reported correlations of .79 and .70, respectively, between patients' and judges' or therapists' descriptions of the patient-therapist relationship.

#### OTHER FACTORS RELATED TO THE PSYCHOTHERAPEUTIC RELATIONSHIP

##### *Patient Variables*

*Diagnosis.* Parloff (1961) found no correlation between initial patient evaluations and subsequent measures of the quality of the therapeutic relationship. Since the patients were all neurotic, the range of maladjustment probably was not very great. Hollingshead and Redlich (1958) reported an inverse relationship between patient maladjustment and therapist liking for patients ranging over the whole gamut of disturbance. However, since maladjustment was also inversely related to social class, and since liking was positively related to social class, the results may be an artifact. Snyder (1961) indicated that the patients toward whom he felt most friendly were relatively more energetic, uninhibited, and suggestible. On the pretherapy MMPI, these patients were more hypomanic and obsessive; the least-liked patients tended to be more schizoid and depressive.

*Prognosis.* Since diagnosis and prognosis are usually themselves related, it should follow that prognosis is correlated with therapists' affective attitudes. Strupp (1958, 1960) and Wallach and Strupp (1960) found this to be true. Strupp (1960) commented on "the ubiquitous but insufficiently realized effects of the therapists' attitudes as they permeate

and color his clinical observations and judgments [p. 28]." Of course it must be realized that the direction of influence may also be the other way.

*Motivation for Therapy.* In three studies (Raskin, 1961; Strupp, 1960; Wallach & Strupp, 1960), therapists had greater liking for patients who evidenced a desire to change and improve. This is hardly surprising since motivation for therapy is usually a central consideration in making a prognosis. Wallach and Strupp inferred from their data that the patient's motivation probably influences the therapist's attitude only when the degree of maladjustment is not too great. Raskin (1961) noted that liking was more effective as a predictor of high- than of low-motivation ratings. Considering the correlational design of his research, the use of the word "predictor" seems inappropriate. Still the finding warrants further study.

*Capacity for Friendliness.* In two studies (Heller, Myers, & Kline, 1962; Snyder, 1961) therapists had warmer feelings for patients who were themselves more friendly.

*Dependency.* Heller et al. (1962) also noted that dependent clients evoked more interviewer friendliness than did more dominant clients. It seems that, at least for therapists judging clients at the outset of therapy, friendliness and dependency are interdependent variables. Presumably this would not hold true at the later stages of therapy.

*Ethical Values.* Snyder (1961) developed a warmer relationship with patients whom he rated as relatively more idealistic and altruistic.

*Social Class.* Hollingshead and Redlich (1958) observed that therapists' attitudes toward their patients were positively related to the patients' social class. This was true even when the range of maladjustment was curtailed by studying only the neurotic sample. The authors

felt that therapists generally are unable to understand lower-class values and hence are less prone to like persons holding such values. Furthermore, the therapist's technical skill is threatened by the lower-class patient's tendency to demand that the therapist behave in an authoritarian manner.

*Expectations Regarding the Therapist.* Apfelbaum (1958) found that patients who expected a high degree of warmth from the therapist did not differ on pre-therapy MMPI scores from patients who expected a low degree, and that both these groups had more maladjusted MMPI scores than a group which expected a moderate degree of warmth. It could be hypothesized that therapists would have the most positive feelings for this middle group.

That therapist liking results from perceiving a "good patient" rather than the reverse was indicated by Strupp and Williams (1960) who found that two therapists' ratings of the following variables were all highly intercorrelated: patients' defensiveness, patients' capacity for insight, patients' motivation for therapy, patients' prognosis, and therapists' liking for the patients. When therapist liking was partialled out, there was no change in the magnitude of the other coefficients. Two writers (Fiedler, 1953; Truax, 1961b) reported contradictory data. In each of these studies, a given therapist's attitudes were found not to vary across different patients. Homogeneity of sampling may account for the null results in both cases.

While it appears that warm relationships can best be established with "good" patients who are motivated for therapy, only moderately maladjusted, friendly, submissive, and who represent the middle or upper social classes, persons who do not fit into this category can sometimes be treated successfully. In order to raise this success rate, it would proba-

ably be fruitful to investigate which sorts of therapists are most capable of forming positive relationships with persons who are not good patients.

### *Therapist Variables*

*Personality Characteristics.* Three studies (Ashby, Ford, Guernsey, & Guernsey, 1957; Brams, 1961; Fiedler & Senior, 1952) correlated a large number of therapist personality variables with measures of the quality of the therapeutic relationship. Positive results occurred no more often than would be expected by chance.

Holt and Luborsky (1952) reported that their successful psychiatrists—who maintained better patient-therapist relationships—were in the middle of the range on dimensions of control of affect and emotional impulsivity. Unsuccessful psychiatrists tended to be either impulsively expressive with their patients or overcontrolled with no expression of feeling. It may be that other therapist qualities have a curvilinear relationship with attitudinal variables and that their relevance has been refuted because only linear relationships were investigated.

Streitfeld (1959) found that ratings of self-acceptance (SA) and acceptance of others (AO) were not related to ratings of therapeutic competence. He concluded that, while general acceptance of others is not related to psychotherapeutic ability, acceptance of specific patients probably is related. The null results might also be explained by the fact that the therapists were all graduate students. Interestingly, therapists who obtained low scores on SA and AO rated their therapeutic ability similarly to the way in which their supervisors rated it, whereas this was not true for therapists high on SA and AO. The author suggested that high SA and AO therapists may be defensive and immature. A similar conclusion may be drawn from data reported by Fiedler and Senior (1952).

Correlational techniques which are sensitive to curvilinearity may be most appropriate for studying the self-acceptance variable.

*Familiarity with the Patient.* Stoler (1961) found that raters who were familiar with patients' cases judged the patients as more likable than did raters reviewing the cases for the first time.

*Ability to Predict Patient Behavior.* A number of studies (Affleck & Garfield, 1961; Chance, 1959; Dymond, 1953; Fiedler, 1953; Luft, 1950; Melton, 1952) have found therapists' predictive ability to be quite poor and generally no better than that of persons who are not therapists. Correlations of this ability, such as it is, with relationship factors have variously been described as positive (Kahn, 1957), negative (Smith, 1960), and zero (Mellinger, 1956). Snyder (1961) reported positive but nonsignificant differences in his ability to predict questionnaire responses of his better and poorer clients.

Fiedler (1953) commented that therapeutic understanding refers primarily to making the patient feel understood and not to any objective diagnostic or predictive ability. The evidence is generally in support of his implication, namely, that predictive ability is not a relevant therapist variable.

*Experience.* Parloff (1956) reported that, for two experienced therapists, judges agreed that one created better therapeutic relationships than the other. Fiedler (1950a, 1951a) found that expert therapists were judged better able to approximate the ideal relationship than were novices. A major weakness in the study was that the judges could distinguish the experts—who were nationally known—from the novices on the basis of their voices. It is impossible to estimate the extent to which halo and "pitchfork" effects, respectively, were operating. Three other studies (Chance, 1959; McGowan, 1954; Strupp, 1958),

in which this defect was avoided, obtained similar results. Hollingshead and Redlich (1958) found that inexperienced therapists tended more to dislike lower-class patients than did experienced therapists. It may be that the experience variable is most relevant in the case of patients who do not fit into the "good patient" category.

### *Therapist-Patient Similarity*

*Assumed Similarity (AS).* This construct (Fiedler, 1951b) refers to the degree to which a therapist believes that his own personality is similar to his patient's personality. Hunt, Ewing, La-Forge, and Gilbert (1959) reported a positive correlation between AS scores and the degree to which therapists liked their clients.

*Real Similarity.* Lesser (1961) found that similarity of self-concept between therapist and patient was not related to empathy scores. Vogel (1961) reported no correlation between judged quality of the therapeutic relationship and patient-therapist similarity on the California F Scale. However, there was a significant correlation between relationship ratings and similarity between patient and therapist ratings of the extent to which the therapist should behave in an authoritarian manner. It should be noted that, in this case, similar opinions would reflect complementary need patterns. Snyder (1961) also reported that patients with whom he had better relationships had need profiles (Edwards Personal Preference Schedule) which complemented, rather than reflected, his own. In this research also, *n* Dominance was the primary factor. Hollingshead and Redlich (1958) noted that therapists had more positive feelings toward patients whose social-class backgrounds were similar to their own.

Carson and Heine (1962) found a curvilinear relationship between therapist-patient similarity on the MMPI and

therapeutic success ratings. That is, patients whose profiles were moderately similar to their therapists' were judged most improved. The authors reasoned that high similarity prevents the therapist from maintaining suitable distance and objectivity whereas high dissimilarity prevents him from being able to empathize with or understand his patient's problems.

These studies strongly support Levinson's (1961) thesis that patient-therapist similarity cannot be thought of as a unitary trait. Some similarities may facilitate good relationships and therapeutic progress, while others may be sources of impasses.

### *Technique Variables*

*Relative Importance of Relationship and Technique.* Shortly before relationship variables came into vogue, Seeman (1949) studied reactions of clients who had been counseled by directive and nondirective methods. He found significantly different client reactions for pairs of counselors whose techniques were similar and nonsignificant differences where techniques were different. He concluded that some factor other than therapeutic method was producing differences in client reactions.

Fiedler (1950a, 1951a, 1953) rushed in to fill the vacuum, affirming Seeman's (1949) conclusion and stating that the therapeutic relationship is the critical variable in successful therapy. Fiedler (1953) commented, however, that a given technique might serve to make a therapist feel more secure and thus indirectly affect the course of therapy.

More recent research (Ford, 1956; . Forgy & Black, 1954; Snyder, 1957) has placed technique in a somewhat more favorable light. These writers have concluded that relationship and success vary not as a function of technique alone, but rather as a function of a particular therapist using a particular technique. They

feel that the therapist and his method must be viewed as a single unit.

Strupp (1958) concurred that both relationship and technique factors are important in successful therapy. Interestingly, in a later article (Strupp, 1962), he was not so willing to commit himself and stated:

For some years it has been held that the quality of the therapeutic relationship is more basic to therapeutic success than the therapist's specific methods and techniques. . . . It may be predicted that the issue will remain alive for some time [p. 452].

*Covariance of Relationship and Technique Factors.* Two studies (Aronson, 1953; Strupp, 1958) reported a negative correlation between the degree of warmth in therapists' attitudes toward their patients and the extent to which they used directive techniques. Apfelbaum (1958) studied patients' pretherapy expectations of their therapists' behavior and reported a curvilinear relationship between the two variables. Therapists expected to be at the extremes of the warmth dimension were expected to employ directive techniques; therapists expected to exhibit a moderate amount of warmth were expected to be relatively more nondirective. The discrepancy between these data and those of Strupp (1958) and Aronson (1953) may be explained in two ways: either therapists do not exist who are as warm as some patients expect or such therapists do exist but were not represented in Strupp's and Aronson's samples. Of course variations in measuring instruments may also account for the differences.

Throughout this paper, considerable weight has been given to methodological issues. It is indeed a pity that research findings must so often be questioned because of methodological defects. Not only the nature of the measurements but also the point at which they are made should be considered, especially in the

case of patients' and therapists' self-reports. While time may dim the memory, it may also extinguish inhibitions and thus allow greater honesty as regards negative feelings and evaluations (Feifel & Eells, 1962).

## REFERENCES

- AFFLECK, D. C., & GARFIELD, S. L. Predictive judgments of therapists and duration of stay in psychotherapy. *J. clin. Psychol.*, 1961, 17, 134-137.
- ANDERSON, R. P., & ANDERSON, G. V. The development of an instrument for measuring rapport. In *Occasional papers in testing and guidance*. Austin: University of Texas, Testing and Guidance Bureau, 1954.
- APFELBAUM, B. *Dimensions of transference in psychotherapy*. (Publications in Personality Assessment and Research, No. 2) Berkeley: University of California Publications, 1958.
- ARBUCKLE, D. S. Client perception of counselor personality. *J. counsel. Psychol.*, 1956, 3, 93-96.
- ARONSON, M. A study of the relationships between certain counselor and client characteristics in client-centered therapy. In W. U. Snyder (Ed.), *Group report of a program of research in psychotherapy*. State College: Pennsylvania State College, 1953. Pp. 39-54.
- ASHBY, J. D., FORD, D. H., GUERNEY, B. G., JR., & GUERNEY, LOUISE F. Effects on clients of a reflective and a leading type of psychotherapy. *Psychol. Monogr.*, 1957, 71(24, Whole No. 453).
- BEHAR, LENORE, & ALTROCCHI, J. Agreement on the concept of the ideal therapist as a function of experience. *J. clin. Psychol.*, 1961, 17, 66-69.
- BRAMS, J. M. Counselor characteristics and effective communication in counseling. *J. counsel. Psychol.*, 1961, 8, 25-30.
- BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers and Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 55-75.
- CARSON, R. C., & HEINE, R. W. Similarity and success in therapeutic dyads. *J. consult. Psychol.*, 1962, 26, 38-43.
- CHANCE, ERIKA. *Families in treatment*. New York: Basic Books, 1959.
- CHASE, W. P. Measurement of attitudes toward counseling. *Educ. psychol. Measmt.*, 1946, 6, 467-473.

- COONS, W. H. Interaction and insight in group psychotherapy. *Canad. J. Psychol.*, 1957, 11, 1-8.
- DYMOND, ROSALIND F. Can clinicians predict individual behavior? *J. Pers.*, 1953, 22, 151-161.
- FEIFFEL, H., & EELLS, JANET. Patients and therapists assess the same psychotherapy. Paper read at American Psychological Association, St. Louis, September 1962.
- FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective, and Adlerian therapy. *J. consult. Psychol.*, 1950, 14, 436-445. (a)
- FIEDLER, F. E. The concept of an ideal therapeutic relationship. *J. consult. Psychol.*, 1950, 14, 239-245. (b)
- FIEDLER, F. E. Factor analyses of psychoanalytic, nondirective, and Adlerian therapeutic relationships. *J. consult. Psychol.*, 1951, 15, 32-38. (a)
- FIEDLER, F. E. A method of objective quantification of certain countertransference attitudes. *J. clin. Psychol.*, 1951, 7, 101-107. (b)
- FIEDLER, F. E. Quantitative studies on the role of therapists' feelings toward their patients. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald Press, 1953. Pp. 296-315.
- FIEDLER, F. E., & SENIOR, K. An exploratory study of unconscious feeling reactions in 15 patient-therapist pairs. *J. abnorm. soc. Psychol.*, 1952, 47, 446-453.
- FORD, D. H. An experimental comparison of the relationship between client and therapist in a reflective and a leading type of psychotherapy. Unpublished doctoral dissertation, Pennsylvania State University, 1956. (*Dissert. Abstr.*, 1956, 16, 1490-1491.)
- FORGY, E. W., & BLACK, J. D. A follow-up after three years of clients counseled by two methods. *J. counsel. Psychol.*, 1954, 1, 1-8.
- GORLOW, L., HOCH, E. L., & TELSCHOW, E. F. *The nature of nondirective group psychotherapy: An experimental investigation*. New York: Teachers College, Columbia University, 1952.
- GRIGG, A. E., & GOODSTEIN, L. D. The use of clients as judges of the counselor's performance. *J. counsel. Psychol.*, 1957, 4, 31-36.
- HELLER, K., MYERS, R. A., & KLINE, LINDA V. Interviewer behavior as a function of standardized client roles. Paper read at American Psychological Association, St. Louis, September 1962.
- HILER, E. W. An analysis of patient-therapist compatibility. *J. consult. Psychol.*, 1958, 22, 341-347.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- HOLT, R. R., & LUBORSKY, L. Research in the selection of psychiatrists: A second interim report. *Bull. Menninger Clin.*, 1952, 16, 125-135.
- HUNT, J. McV., EWING, T. N., LAFORGE, R., & GILBERT, W. M. An integrated approach to research on therapeutic counseling with samples of results. *J. counsel. Psychol.*, 1959, 6, 46-54.
- KAHN, R. K. Therapist discomfort in two psychotherapies. Unpublished doctoral dissertation, Pennsylvania State University, 1957. Cited by W. U. Snyder, Some investigations of relationship in psychotherapy. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington D. C.: National Publishing, 1959. P. 252.
- KNUPFER, GENEVIEVE, JACKSON, D. D., & KRIEGER, G. Personality differences between more and less competent psychotherapists as a function of criteria of competence. *J. nerv. ment. Dis.*, 1959, 129, 375-384.
- LESSER, W. M. The relationship between counseling progress and empathic understanding. *J. counsel. Psychol.*, 1961, 8, 330-336.
- LEVINSON, D. J. The psychotherapist's contribution to the patient's treatment career. Paper read at Conference on Psychotherapy Research, Chapel Hill, North Carolina, 1961.
- LIPKIN, S. The client evaluates nondirective psychotherapy. *J. consult. Psychol.*, 1948, 12, 137-146.
- LIPKIN, S. Clients' feelings and attitudes in relation to the outcome of client-centered therapy. *Psychol. Monogr.*, 1954, 68(1, Whole No. 372).
- LUBORSKY, L. B. The personalities of successful and less successful psychotherapists. *Amer. Psychologist*, 1952, 7, 337. (Abstract)
- LUFT, J. Implicit hypotheses and clinical predictions. *J. abnorm. soc. Psychol.*, 1950, 45, 756-759.
- MCCLELLAND, W. A., & SINAICO, H. W. An investigation of a counselor attitude questionnaire. *Educ. psychol. Measmt.*, 1950, 10, 128-133.
- McGOWAN, J. F. Client anticipations and expectancies as related to initial interview performance and perceptions. Unpublished doctoral dissertation, University of Missouri, 1954. (*Dissert. Abstr.*, 1955, 15, 228-229.)
- MELLINGER, G. D. Interpersonal trust as a factor in communication. *J. abnorm. soc. Psychol.*, 1956, 52, 304-309.
- MELTON, R. S. A comparison of clinical and actuarial methods of prediction with an assessment of the relative accuracy of different

- clinicians. Unpublished doctoral dissertation, University of Minnesota, 1952.
- MILLER, HELEN E. "Acceptance" and related attributes as demonstrated in psychotherapeutic interviews. *J. clin. Psychol.*, 1949, 5, 83-87.
- PARLOFF, M. B. Some factors affecting the quality of therapeutic relationships. *J. abnorm. soc. Psychol.*, 1956, 52, 5-10.
- PARLOFF, M. B. Therapist-patient relationships and outcome of psychotherapy. *J. consult. Psychol.*, 1961, 25, 29-38.
- RASKIN, A. Factors therapists associate with motivation to enter psychotherapy. *J. clin. Psychol.*, 1961, 17, 62-65.
- ROGERS, C. R. An overview of the research and some questions for the future. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univ. Chicago Press, 1954. Pp. 413-430.
- ROGERS, C. R. The necessary and sufficient conditions of therapeutic personality change. *J. consult. Psychol.*, 1957, 21, 95-103.
- ROGERS, C. R. Therapist-patient relationship. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: National Publishing, 1959. P. 273.
- SEEMAN, J. An investigation of client reactions to vocational counseling. *J. consult. Psychol.*, 1949, 13, 95-104.
- SEEMAN, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univ. Chicago Press, 1954. Pp. 99-108.
- SHOBEN, E. J., JR. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, 46, 366-392.
- SHOBEN, E. J., JR. Some observations on psychotherapy and the learning process. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald Press, 1953. Pp. 120-139.
- SMITH, W. Social adjustment and interpersonal perception. In J. McV. Hunt (Ed.), *Interpersonal perception and interpersonal relationships in therapeutic counseling*. Technical Report No. 1, 1960, University of Illinois, ONR Contract Nonr-1834 (11), Office of Naval Research.
- SNYDER, W. U. The psychotherapy research program at the Pennsylvania State University. *J. counsel. Psychol.*, 1957, 4, 9-14.
- SNYDER, W. U. *The psychotherapy relationship*. New York: Macmillan, 1961.
- SOPER, D. W., & COMBS, A. W. The helping relationship as seen by teachers and therapists. *J. consult. Psychol.*, 1962, 26, 288.
- STOLER, N. Client likability as a variable in the study of psychotherapy. *U. Wisc. Psychiat. Inst. Bull.*, 1961, 1(No. 11).
- STREITFELD, J. W. Expressed acceptance of self and others by psychotherapists. *J. consult. Psychol.*, 1959, 23, 435-441.
- STRUPP, H. H. The psychotherapist's contribution to the treatment process. *Behav. Sci.*, 1958, 3, 34-67.
- STRUPP, H. H. *Psychotherapists in action*. New York: Grune & Stratton, 1960.
- STRUPP, H. H. Psychotherapy. *Annu. Rev. Psychol.*, 1962, 13, 445-478.
- STRUPP, H. H., & WILLIAMS, J. V. Some determinants of clinical evaluations of different psychiatrists. *Arch. gen. Psychiat.*, 1960, 2, 434-440.
- SUNDLAND, D. H., & BARKER, E. N. The orientation of psychotherapists. *J. consult. Psychol.*, 1962, 26, 201-212.
- THOMAS, E., POLANSKY, N., & KOUNIN, J. The expected behavior of a potentially helpful person. *Hum. Relat.*, 1955, 8, 165-174.
- TRUAX, C. B. The process of group psychotherapy: Relationships between hypothesized therapeutic conditions and interpersonal exploration. *Psychol. Monogr.*, 1961, 75(7, Whole No. 511). (a)
- TRUAX, C. B. Therapeutic conditions. *U. Wisc. Psychiat. Inst. Bull.*, 1961, 1(No. 10c). (b)
- TRUAX, C. B. Constructive personality change in schizophrenic patients receiving high-condition therapy, low-condition therapy, and no therapy. Unpublished manuscript, University of Wisconsin, Wisconsin Psychiatric Institute, 1962.
- VAN DER VEEN, F. The perception by clients and by judges of the conditions offered by the therapist in the therapy relationship. *U. Wisc. Psychiat. Inst. Bull.*, 1961, 1(No. 10).
- VOGEL, J. L. Authoritarianism in the therapeutic relationship. *J. consult. Psychol.*, 1961, 25, 102-108.
- WALLACH, M. S., & STRUPP, H. H. Psychotherapists' clinical judgments and attitudes toward patients. *J. consult. Psychol.*, 1960, 24, 316-323.

(Received April 18, 1963)

# NEED STATES AND ACTIVITY LEVEL

ALFRED BAUMEISTER  
*Central Michigan University*

WILLIAM F. HAWKINS AND RUE L. CROMWELL<sup>1</sup>  
*George Peabody College*

Increases in activity level have often been used to imply the presence of a heightened drive state. Frequently, drive is also anchored on the antecedent side to a condition of need. Research literature which bears on the relationship between need state and activity level is reviewed. Specifically, activity level is discussed in terms of food, water, sex, and activity deprivation. It is suggested that associative interpretations of activity level may be more meaningful than drive interpretations.

Students of motivation who employ the concept of "drive" as the origin or source of the behavioral activity of an organism do not, in their experimental investigations, actually identify or measure the specific factors in the organism which constitute the drive. Hence, drives are unknowns, postulational or inferential in character under our present state of knowledge. Their spatial locus is the organism, their temporal is in the period immediately preceding the neural innervation which results in motility or activity.

The term "drive," drawn from general vocabulary sources, has implications beyond mere antecedence—its connotation is similar to "forced," in Loeb's classic definition of tropisms. But it is the amount of activity, rather than its direction or orientation, or even the character of the responses, which investigators of physiological drives have usually sought to measure and to correlate with the known and identifiable circumstances which operate prior to the postulated drives. The behavior of organisms which is usually measured may be primarily locomotor, or it may consist of postural shifts, or be even more vaguely

described as "restlessness," or termed "general," "random," or "spontaneous" activity.

The precursor or antecedent of drive has been designated by some as a need state. For the purposes of this discussion, "need state" may be defined as any condition of deficit or excess which causes the organism to depart from the dynamic equilibrium known as homeostasis.

Research upon the relationships between need, drive, and activity encounters many and varied problems which do not admit of any easy or entirely satisfactory solutions. Among these the following may be cited as fundamental:

1. The determination of the base line of activity since the organisms studied all manifest some motility even under homeostasis (or what one is tempted to call a "steady state"), from which changes in level are to be evaluated.

2. The quantitative control of the conditions (need states) which result in a significant change in activity from that which characterizes or constitutes the base line.

3. The measurement of the activity level.

4. The specification of the processes by which need states are translated into overt activity. The most common approach here is to postulate an interven-

<sup>1</sup> The authors would like to express their gratitude to K. C. Pratt and N. R. Ellis for their helpful suggestions in the preparation of this manuscript.

ing drive state. Changes in activity level which are correlated with changes in need state are frequently attributed to the "energizing" function of drive (Bolles, 1958).

These are not simple problems. Indeed there are grounds for questioning whether any truly unitary constructs are warranted at this time. Activity level at present defies any definition except in terms of measurements obtained from various activity wheels, stabilimeters, photoelectric devices, open-field tests, etc. As we shall see, the data obtained with one type of measuring device are not necessarily the same as those from another device under the same need states.

The aim of this paper is to summarize and to evaluate some of the experimental literature in which there has been an attempt to relate activity level to need state.

#### FOOD RESTRICTION

The most intensively and carefully studied need states, as they relate to activity, are hunger and thirst. Common observation and experience tell one that as an animal becomes hungry or thirsty it behaves in a manner that typically relieves this condition; that is, it reduces the physiological deficit and returns to a state of relative equilibrium. The animal will run about, pacing, sniffing, and in general will appear to explore its environment. Observation of infant animals indicates that they display considerable random activity even after short periods of food and water deprivation. Obviously, an organism must show some mobility (at least in its natural habitat) in order to survive. Research findings tend to support informal and common-sense observations that rate of response is related to severity of the deprivation. Activity, thus,

has become a measure of drive. Or, stated more appositively, heightened activity level in the presence of an altered need state has been attributed to the energizing function of drive.

Many investigators have shown that activity (mainly in a running wheel) increases up to a point as a function of food and water deprivation. One of the clearest examples of this relationship is found in the work of Moskowitz (1959), who reports that an almost linear function obtains between running wheel activity in the rat and decrease in body weight as controlled by food and water deprivation. Other studies (e.g., Finger, 1951; Hall, 1956) have shown that activity in wheels increases directly as a function of the number of hours that rats are deprived of food. An increase in activity in relation to food deprivation has sometimes been noted with respect to bar pressing (Segal, 1959; Webb, 1948), beam interruptions (Siegel & Steinberg, 1949), open-field behavior (Dashiell, 1925; Fehrer, 1956), treadmill running (Baba, 1959), and stabilimeter behavior (Richter, 1922; Teghtsoonian & Campbell, 1960; Teitelbaum, 1957). Irwin (1932), using the stabilimeter technique, found a significant correlation between the activity of human infants and lapse of time from a feeding period. Among the more well-established findings is the fact that animals show regular cycles of activity and eating. Following eating, activity decreases markedly (Finger, 1949; Richter, 1927).

Unfortunately, many qualifications cloud the apparent general relationship between deprivation and activity. Among the considerations that limit this generalization are: (a) findings that the respective deprivation states do not consistently produce similar changes in activity; (b) that intercorrelations between measurements from

the various devices are generally low; (c) that the relationship may be an indirect one which is, in turn, dependent upon stimulus conditions, learning, and physical condition; (d) that different species, perhaps even strains, show different effects; (e) that sex differences are apparent in activity level; and (f) that age is a contributing factor. Then, too, the interactions among these factors must be considered. For example, it may be that changes in activity level as a function of food deprivation are dependent upon the nature of the apparatus and the stimulating conditions.

The question arises whether deprivation *per se*, as manifested by changes in blood chemistry or some other physiological mechanism, is a direct antecedent factor in increasing activity. That is, are increases in activity level the unlearned effects of the physiological consequences of deprivation, or does deprivation sensitize the organism to react to features and changes in its environment to which, in a satiated state, it does not respond? According to the latter point of view one might expect activity to remain relatively stable in an unchanging environment even though the organism were deprived. Moreover, the activity should increase as a function of change in the environment and increased need state. This is a fundamental issue. It raises the question whether or not activity adequately and directly reflects the strength of drive.

Campbell and Sheffield (1953) contested the classical view that activity is a resultant of an increased drive state due to a heightened condition of primary need. They found only minimal stabilimeter increments in activity over 72-hour food-deprivation intervals when external factors related to vision and audition were held constant by rigid environmental control. They explained their findings by suggesting that hunger serves to reduce perceptual thresholds

for external stimuli and that it is these stimuli, in turn, which initiate activity. Segal (1959) studied bar-pressing behavior in relation to these factors and came to much the same conclusions. This position bears a similarity to one advanced by Geier (1942), who contended that the tension produced by the expectancy of food makes the organism more sensitive to the stimuli impinging upon it. In effect, these writers reject the traditional notion of drive as an energizer and assign to it the role of alerter.

Several additional studies offer support for the hypothesis that the primary condition leading to increases in deprivation-related activity is external stimulation, while conditions of need merely serve to sensitize the organism to stimulation. Baker (1953), in examining closely the cyclic nature of activity, has found that the characteristic cycle is absent if the animal is deprived and its environment is held constant. Lubow and Tighe (1957) reported that increases in illumination raise activity level in a stationary apparatus in a degree proportional to the amount of change in intensity. Decreases in illumination, however, appeared to have little effect. Similarly, Roberts (1962) found that stimulus changes produce recovery in bar-pressing behavior. Monkeys showed no alteration in activity level up to 54 hours of food deprivation in a reasonably constant environment (DeVito & Smith, 1959). Teghtsoonian and Campbell (1960) and Amsel and Work (1961) have compared the activity of deprived animals whose environment was held constant with others subjected to changes in the level of stimulation. The effect was the same. The nonstimulated group remained near its base-line activity while the other group showed substantial increases. Nevertheless, in mice, at least, the relationship between changes in illumi-

nation and activity is partly dependent upon strain (McClern, 1959, 1960). It should be pointed out, however, that studies of human infants (Irwin, 1930, 1941) suggest that their activity decreases as illumination increases. Periods of continuous auditory stimulation also show decreasing activity as compared with periods which contain no auditory stimuli. Further, if the infant is subjected to both illumination and auditory stimulation there is an even greater reduction in stabilimetric activity—in other words, the presence of these forms of energy are additive in reducing activity.

It should be pointed out in regard to these studies that the term "constant" can only be applied in a relative sense. That is, the control animals by virtue of moving about in their cages, however minimal such movement may be, are changing the environment to the extent that the stimulus conditions are altered. Furthermore, it is apparent that in spite of the most rigorous efforts of the experimenter to maintain a constant environment, small variations in noise, temperature, odors, etc., are bound to occur.

Campbell and Sheffield (1953) further theorize that activity is a function of the stimulus effect of fractional anticipatory eating responses which are conditioned to the external environment. While it is difficult to comprehend the exact process by which anticipatory eating responses would be conditioned to environmental changes, particularly those which the animal has not experienced previously, or which it cannot consistently associate with hunger reduction, it does seem that prolonged deprivation of specific foods might bring about a lowering of thresholds for specific kinds of stimulation or that general deprivation might be followed by an increased sensitivity to almost any variety of stimulation.

Another possible interpretation is suggested by the arousal function of the reticular formation (Bolles, 1958). Centers that respond to specific deficits may exert a disinhibiting influence on the ascending reticular activating system, thereby leaving the animal more responsive to external stimulation. However, there is not necessarily a one-to-one relation between events occurring within the central nervous system and overt behavior. Therefore, while arousal mechanisms within the central nervous system might provide a useful model on which to base a concept of drive, drive and arousal should not be considered as interchangeable constructs, particularly since many of the antecedent conditions which define drive state do not seem to be reflected in changes in cerebral activity.

In an effort to examine more closely the role of associative factors with respect to variations in activity, Sheffield and Campbell (1954) introduced controlled changes (an electric fan and lights) into the external environment prior to regular feeding. Greater activity was manifested by rats fed at the end of the stimulus period than by similarly deprived controls that were fed with no preceding environmental change. They interpreted this increment in activity as an effect of frustration. Such frustration was produced when the conditioned external changes occurred and when the goal, and consequently the consummatory response, was not possible. The stronger the connection between consummatory responses and change, the more frustrated (active) the organism will be when the changes occur and when there is no possibility of the consummatory response. Amsel and Work (1961) repeated the experiment by Sheffield and Campbell, but included another group of animals which were fed on half of the stimulus-change trials. Partial reinforcement of

activity, while producing changes, does not result in the sharply ascending activity curve which characterizes the 100% group. Amsel and Work suggested that the 50% group did not have a sufficient number of trials to develop the  $r_g$  mechanism accounting for anticipatory goal reactions. Thus, the frustrating effects of nonreward were not as great in this latter group.

But, it is not clear whether the frustration affects activity by directly increasing drive, or whether various motor responses are learned in connection with frustration-produced proprioceptive stimuli. One is inclined to wonder just how the notion of frustration-produced activity relates to the concept of sensitization, unless one accepts the implicit suggestion that deprivation sensitizes the organism to various cues which are frustrating.

In the interests of parsimony, it might be suggested that this phenomenon can be explained without reference to states of "frustration" or of "tension." Some activity inevitably precedes eating and it is suggested that this activity is reinforced by consummatory responses. By virtue of feeding the animal, its mother or the experimenter consistently establishes changes in the environment. This nonspecific stimulation (e.g., sound, touch, odor, etc.) becomes a conditioned stimulus which, after a sufficient number of pairings with the unconditioned stimulus of food, can elicit the activity response of approaching the food. Such associations would seem to occur very early in the animal's experience. Furthermore, activities more and more remote in time, and consequently in adaptiveness, are linked together through a process of higher order conditioning and generalization. This should produce a chain of conditioned responses, as a function of the reinforcement contingencies and magnitude of deprivation, which would

gradually build up to reach its peak in magnitude just before reinforcement.

The total stimulus complex to which the activity response is conditioned consists of two sets of cues: those provided by the external environment and those which result from physiological mechanisms activated by the need state. For example, the introceptive cues associated with water deprivation would arise from stimulation in the pharyngeal regions, the stomach, and the hypothalamus. The internal stimuli associated with a given level of need differ from those of any other level of need, both with respect to number and intensity. The greater the need state, the more intense and numerous the introceptive cues.

Thus, changes in activity may be a function of the degree to which the total maximal stimulus complex is represented at a given moment by the available cues. It would follow from these notions that the general activity level should be raised by changes in the external environment, but that deprivation should increase it even more. This, of course, is not what Campbell and Sheffield have suggested. Nevertheless, they do report significant increases in activity in the nonstimulated group. Another study (Hall, 1956) clearly substantiates our hypothesis. One might also predict that the shape of activity curves plotted as a function of length of food deprivation would be different for stimulated and nonstimulated groups. Because there are fewer maintaining stimuli in the latter condition, deceleration should occur more rapidly and an asymptote reached sooner. The extinguishing effects of nonreward of activity should be more rapid in the nonstimulated group. Finally, one might hypothesize that following eating, when the need-related cues are virtually eliminated, activity level would be at its lowest.

This particular phenomenon has frequently been observed.

In some respects these suggestions are similar to those advanced by Estes (1958). Estes, however, refers to drive stimuli rather than to need-related cues.

Finger, Reid, and Weasner (1957, 1960) have demonstrated that manipulation of the reinforcement schedule alone modifies running activity. By feeding one group of subjects immediately after removing them from the wheels and another group 60 minutes later, these workers were able to produce differences not only with respect to total activity but also in terms of its pattern of development. Moreover, this phenomenon occurs under conditions of relatively low need. Hall (1958) has confirmed the finding by Finger and his associates and suggests that activity-wheel behavior is not a suitable indicator of drive strength when a deprivation-feeding cycle is used. Seward and Pereboom (1955) have reported similar results. Apparently, activity-wheel running is subject to many of the same laws as other forms of instrumental behavior. Again, in some of those studies in which exploratory behavior has been related to deprivation states (e.g., Alderstein & Fehrer, 1955), learning has not been ruled out as an explanatory factor. Particularly important here is the schedule of deprivation and ingestion with respect to measurement of exploration. In an attempt to control for the effects of deprivation schedules, Glickman and Jensen (1961) have found that moderate degrees of hunger (22-hour period) do not produce significant changes in exploratory behavior, but that a 46-hour period of deprivation does lead to an increase in maze exploration. It is interesting to note from their data that similar periods of thirst did not result in significant changes in exploratory behaviors.

Even in the instance where there is

suitable experimental control for the reinforcing effects of posttest feeding on activity, increases in activity might still be attributable to learning by virtue of cues specific to a particular degree of privation. Thus, for example, the rat may have learned that a vigorous response to intense need-produced stimuli is more adaptive than a weak response, both in terms of time and amount. Jenkins and Hanratty (1949) have shown that rats can learn to respond differently to different degrees of deprivation. In short, it is probably safe to say that the rats' early experience with deprivation and feeding have provided that basis for learning to be active in response to cues associated with deprivation.

In any event, Campbell and Sheffield's (1953) interpretation could be an important one for a theoretical appraisal of activity since it conceives of activity as dependent upon external factors and learning as well as upon need states. A further and perhaps more important implication is that activity can be viewed as directional, however inappropriate the response may appear to be. Activity, in other words, may not be spontaneous, but rather generalized approach or avoidance behavior. It is possibly best to regard such behavior as instrumental. If one holds to a drive interpretation of activity, then in terms of the results of the investigations of Campbell and his associates it would appear that the role of drive is merely that of an alerting mechanism.

Nevertheless, a portion of the literature, if not directly refuting the sensitization hypothesis, at least fails to confirm it. Hall and Hanford (1954) controlled the external environment by holding light and temperature constant and by deafening their animals. They reported that wheel activity increased as much in the experimental animals as in the control group. Hall, Smith,

Schnitzer, & Hanford (1953) also found a relationship between activity and deprivation independent of external sources of stimulation. The results of a later study by Hall (1956) have already been mentioned. Part of the difficulty in evaluating the literature stems from procedural differences both in terms of the actual measurements and also with respect to the conditions under which they were obtained. Surely, it is fairly well established that a relationship exists between food deprivation and activity level. But, whether this occurs independently of the external environment, of learning, and of other factors, is another, and as yet, unresolved question.

#### FOOD VERSUS WATER RESTRICTION

Most of the literature which we have reviewed has dealt with food deprivation upon activity. What is the situation with respect to restriction of water?

Finger and Reid (1952) found that water restriction affected running activity as much as did food deprivation with performance curves increasing continuously under both conditions up to 72 hours. A similar finding is reported by Stevenson and Rixon (1957). However, the "satiation syndrome," a term coined by Finger (1949) to describe the typical depression of activity following relief of deprivation, is less apparent under conditions of water restriction. This suggests that stimulation associated with digestive activity may be a factor relevant to overt activity.

There is a considerable lack of consistency, however, in the research findings with respect to comparisons of these need states. For instance, Hall (1955) did not observe any increase in running activity under prolonged restriction of drinking. This confirms an earlier finding by Richter (1922). Similarly, Bolles (1959) reported increases in running related to food deprivation

but not with respect to thirst. Stabilimeter measurements indicate, according to Campbell (1960) that stimuli introduced just prior to drinking or eating result in heightened activity level, whereas water deprivation alone produces no noticeable effect. This result is cited as further support for the sensitization-frustration theory.

It should be noted that a complicating factor arises in the employment of either food or water deprivation in that when one of these is restricted, consumption of the other falls below normal intake. Thus a food-deprived animal, in terms of his base-line consumption of water, is also thirsty, and vice versa.

In addition, there is no reason to suppose that a given period of food deprivation is equal in its drive consequences (or in its stimulus value) to a similar period of water deprivation. Dufort and Blick (1962) have shown that rats adjust much more rapidly to water deprivation than to food deprivation. It is clear that there is an abundance of methodological problems present and any general statement regarding the relative effects of food and water deprivation upon activity must be withheld. There is a critical need for research into the mechanisms by which the different need states affect the internal receptors of the organisms in overlapping and in separate ways.

#### SPECIFIC NUTRIENT DEPRIVATION

Specific-nutrient-deprivation studies have been reviewed by Spector and Young (1957). In general, there seems to be an initial increment in activity related to internal and/or external stimulation. This activity leads to further depletion and, if there is no repair, to exhaustion with a decrease in activity (Wald & Jackson, 1944). However, the picture is not universally as simple as this.

Reed (1947) and Munn (1950) have

indicated that activity is related to the percentage of proteins in the diet. Transitory increases in activity seem to follow decreased protein intake.

There are numerous examples of the effects of specifically restricted diets upon the activity of animals. Richter (1957) reported that rats are most active on a diet of carbohydrates and dextrose, less active on fats and lards, and least active on protein and casein. Animals deprived of thiamin and riboflavin, according to Wald and Jackson (1944), show an increase in activity. But when given sufficient amounts of these vitamins they tend to become more quiescent than the controls. Restriction of vitamin B in the diet leads first to increased activity and then, after a few days, to lessened activity. On the other hand, deficiency of either vitamin A or vitamin D appears to result in an almost immediate decrement in activity (Richter & Rice, 1942, 1943).

These divergent findings are troublesome for drive interpretations of activity. While the composition of the diet does seem to influence gross bodily activity, the changes appear to be highly irregular. Moreover, there are some types of dietary deficiencies which result in an immediate decrement in activity and some which have no apparent effect. These findings serve to indicate the complexity of the activating mechanisms. While there is no requirement that drive and need be regarded in a one-to-one relation, neither is it very satisfying to explain the absence of an increase in activity by referring to a particular type of need which does not initiate drive. Perhaps it would be more appropriate to refer to the stimulus properties of the various need states. Indeed, it might be suggested that the internal cues arising from total food deprivation are an accumulation deriving from the numerous specific needs.

Thus, one might suspect that the internal stimuli associated with food deprivation are more numerous, but more variable from moment to moment than those resulting from water deprivation. These conditions may account for the more rapid adaptation to water deprivation as well as the greater increase sometimes reported in activity associated with food deprivation.

#### SEX

Activity level appears to be related to sexual need, as defined by estrous variations. At the peak of the 4-5 day estrous cycle, running activity is several times as great as it is during other portions of the cycle (Finger, 1961b). However, when pregnancy occurs and there are no further periods of receptivity, activity drops to a consistently low level (Brobeck, Wheatland, & Strominger, 1947; Wang, 1923). After sex glands are removed, thus abolishing the female cycle, activity is reduced to a permanently low level (Richter, 1927).

Brody (1942) and Finger (1961b) have shown that, in comparison to males, female rats are four times as active in revolving drums and twice as active in stationary cages. While this difference may be explained, in part, by the increased activity of the females in estrus, the consistency of the difference lends support to the view that other factors are involved. Apparently, this finding cannot be extended to humans. Wolfensberger, Miller, Foshee, and Cromwell (1962) have taken ballistographic measures of activity in junior and senior high school students. They report that boys are more active than girls. Although this sex difference is a common observation in young children, cultural factors may be contributing to the picture here. Schulman, Lipkin, Clarinda, and Mitchell (1961) have obtained essentially the same results

using the actometer as the measuring device. Childers (1935), on the other hand, found no sex or race differences in activity level.

A male rat placed in an activity wheel next to a female will display peaks of activity related to the female's estrous period. When a number of females are placed in his vicinity, the male's activity increases, but it is at a more constant level (Reed, 1947). One explanation of this finding is that the females are in heat at different times and the male responds to all of them. A complicating factor here for exponents of drive-reduction theories is the finding that copulation does not serve to reduce the running-wheel activity of the male (Morgan & Stellar, 1950). Nor does there seem to be much connection between running-drum activity level and the sex drive, as measured by the readiness of the male to copulate with receptive female (Stone & Barker, 1934). In this instance, then, there does not appear to be any relation between drive state and activity, at least as measured by the activity wheel.

As contrasted with the relation of activity to nutrient deprivation, the causative factors are less elusive in the female sex drive to activity. It is quite clear that sexual hormones play a substantial role. Both sexual receptivity and sexually linked activity are probably based on a direct effect to central structures by chemical conditions determined in the ovaries (Morgan & Stellar, 1950). Hormonal conditions, likewise, are probably implicated in sex-related activity of the male.

#### NEED FOR ACTIVITY

Up to this point, we have been discussing activity as an indicator of need. The question arises, however, whether activity can also be a drive. That is, does a condition of restriction of movements for a period of time produce extra

activity when the restriction is removed? The typical research paradigm followed is that of confining rats and studying the effects of the confinement on subsequent activity. If there is such a state as an activity "drive," then the confined animal should display more activity at the end of the confinement period, the level of activity being a function of the length of confinement.

Some investigators have indicated that activity increases following confinement (e.g., Baron, Antonitis, & Beale, 1961), but the effects are not clear-cut. Siegel (1946), for instance, found that 24-hour confinement leads to an increase in activity, but 6- to 12-hour confinements were followed by decreases in activity. Siegel and Alexander (1948), in repeating Siegel's (1946) earlier study, found that the relationship between enforced confinement and activity is dependent on the time of the day. The 6-hour confinement period has no effect if the confinement occurs from early to late morning, a period during which the rat is normally inactive. However, the depressing effect does occur if confinement is imposed between late morning and early evening, when the laboratory rat is relatively more active. Hill (1956, 1958, 1961), attempting to demonstrate the generality of the relationship, has shown that activity in the wheel, in a tilt apparatus, and in number of squares traversed in an open field increases among rats according to the length of confinement within the range of 5 hours to 40 days. The age of the animal is apparently a contributing factor here, there being an interaction between age and length of deprivation. Hill concluded that his results support an activity drive conclusion. On the other hand, it should be added that the studies by Montgomery (1953a, 1953b) led him to suggest that activity deprivation produces no increase in exploratory be-

havior. The weight of the evidence suggests that when animals are not allowed to behave in certain ways, the opportunity to do so produces an increase in such behaviors.

Apparently, behavior is often energized in situations where the contributing antecedent factors are as yet unknown. This is apparent in the play activity of animals (Beach, 1945). Rats will run although their needs seem to be well satisfied. Children are active even when physiological needs appear to be at a minimum. Monkeys evidently will learn with nothing more than manipulation serving as a reinforcement (Harlow, 1954). Exploration may motivate learning (Montgomery, 1954), as does curiosity (Berlyne, 1955). Curiosity, manipulation, and exploratory behaviors are viewed by some as possible special cases of an activity drive (e.g., Morgan, 1957). It should be pointed out that curiosity, manipulation, exploratory, and activity drives are overlapping constructs when they are reduced to operational terms.

Kagan and Berkun (1954) and others have produced further support for a drive interpretation of activity by showing that it can serve as reinforcement for learning. In their study, an activity wheel was designed so that the animal could release brakes on the wheel by pushing a lever. By releasing these brakes, the animal could turn the wheel for a period of 30 seconds until the brakes were reapplied. Animals under these circumstances responded considerably more than did rats in freely moving wheels. Unfortunately for a drive interpretation of activity, it has been shown (Hill, 1961; Montgomery, 1954) that confinement does not lead to a substantially greater preference of an activity incentive.

These experiments indicate that, in spite of the lack of nutrient-based cues or intense sensory stimulation, a need

for activity seems to be manifested. However, it is not clear what role associative factors play. Conceivably, cues arising from confinement or inactivity (need for exercise) may be associated with various types of overt motor responses. For instance, normal circulation of blood is partly dependent upon muscular contraction.

There can be little doubt that the need states resulting from food, water, and sex restriction through their central mechanisms, energize sensory-motor systems. Some of these same structures may be involved in the activity drive. There is considerable literature (Lindsley, 1957) which indicates that the reticular activating system and the hypothalamus are involved in the activity need. Vertebrates with damage in these areas show altered activity levels. Animals with lesions in their "sleep center" show increased activity. Damaged "wake center" animals show lessened activity. Insult to the frontal lobe and in the area of the corpus striatum is associated with increased activity level. The work of Campbell and his associates (Campbell & Sheffield, 1953; Sheffield & Campbell, 1954; Teghtsoonian & Campbell, 1960) suggests the involvement of the ascending reticular activating system as a mediating center.

#### METHODOLOGICAL PROBLEMS

As indicated earlier in this review, a portion of the conflicting evidence might be explained on the basis of differences in methodology. The devices most frequently used in measuring activity are the stabilimeter and the wheel. Each of these has undergone many modifications. Only a limited amount of investigation of the interrelations among various measures of activity has been attempted. The information which is available is sufficient to suggest that activity level should not be viewed as a

single or homogeneous phenomenon. Many writers (e.g., Morgan & Stellar, 1950; Reed, 1947) have suggested several dimensions of activity—locomotor versus diffuse, relevant versus irrelevant, and goal-directed versus non-goal-directed. Reed (1947) comments that "what data we have point to more than one type, or at least more than one aspect, of activity." Hunt and Schlosberg (1939) have made a similar statement.

How do the various measures compare with one another? Some studies (Bolles, 1959; Furchgott & Echols, 1958; Hayes, 1957; Stern, 1957) indicate that open-field behavior is distinct from wheel running and stationary cage activities. Hunt and Schlosberg (1939), studying castration, and Tainter (1943), investigating various drugs, reported that their experimental treatments differentially affected stabilimetric and running-wheel activity. More recently, Strong (1957) has found that food deprivation will lead to an increase in diffuse activity as recorded by a sensitive contact stabilimeter, but if a micro-switch stabilimeter is employed, food-deprived rats do not differ significantly in activity from their controls or in relation to predeprivation level. Along these same lines, Finger (1958) measured activity with both the revolving drum and photoelectric techniques and found that the drum measure was more influenced by increases in food deprivation. In another study, Weasner, Finger, and Reid (1960) have shown that with different deprivation regimens the running increases proportionately more than does activity in a stationary device. Treichler and Hall (1962) have compared activity in the wheel, stabilimeter, and Dashiell Maze as a function of body weight. They concluded that the three measures do not produce interchangeable data, the stabilimeter and maze scores showing little change as a result of weight loss.

The picture is much the same when need variables other than food restriction are used. For instance, Finger (1961b) has demonstrated that wheel activity is more responsive to estrous variations than stationary cage activity. A further complication is introduced by the finding that the difference between these two measures, as a function of estrus, is not as great as the disparity noted when the animals are subjected to food deprivation. In other words, we may infer a triple-order interaction between need states, degree of privation, and measuring device.

Strong (1957) has claimed that the stabilimeter type of measurement used by Campbell and Sheffield (1953) detects and records relatively more diffuse activity of fine movements than some of the methods employed by others. Hunger appears to increase locomotor activity as measured by the wheel and photoelectric cages, but seems to have considerably less effect on stabilimeter activity. On the other hand, it may be that external stimulation increases restless behavior to which the stabilimeter is more sensitive. Thus, it seems that the conflicting results of Campbell and Sheffield (1953) and Hall (1956) are explicable on the basis of differences in the recording of activity changes. In this connection, it may be argued that wheel rotations give a superior method of quantification due to the inherent specificity of response. Two units of increase, for example, in an activity wheel allow one to assume that their combined "activity" is approximately twice as much as a 1-unit increase. However, in the stabilimeter this assumption is not nearly as tenable because the score producing response acts themselves are not truly additive. This argument would apply to the photoelectric devices as well.

Another feature which should be noted is the relative similarity of the

stabilimeter to the home environment of the typical laboratory rat. The responses which the rat is physically able to make are virtually identical in the two situations. However, the wheel makes it possible for the animal's behavior to take more complete and less restricted locomotive forms, rather than limiting movements to postural responses. This difference may very well account for the finding (Eayrs, 1954) that habituation to the stationary cage occurs much more rapidly than adaptation to the revolving drum.

A related methodological problem is the determination of the base line for the activity of a given animal. By first measuring the normal activity of the organism it is possible to estimate more precisely the effect of a particular treatment. The base line is usually considered established when there is no consistent change over a period of time, independent of the normal cyclic behavior and aging process. When the habituation requires an extended period, this latter factor might be important, although it is rarely considered.

There is considerable interexperimenter variability with respect to what is considered necessary to establish a suitable base line. Campbell and Sheffield (1953), for example, used much shorter intervals (10 minutes a day for 7 days) than most other researchers. Finger (1961a) and Reid and Finger (1955) have demonstrated that base-line activity in the running drum takes at least 30-35 days to reach a stable level. Moreover, the cyclic feeding experiences of the rat before it has ever come into contact with the wheel complicate the matter. If the animal is placed on cyclic deprivation prior to wheel running, the base-line function becomes fairly level within 15 days (Reid & Finger, 1957). The time required to establish a reliable base line partially dependent upon the

apparatus employed—stabilimetric-type cages requiring only a few days. In most studies, at least a portion of the variance in activity scores is probably due to habituation factors.

Still another factor which must be considered in an attempt to compare the different experiments relates to differences in deprivation schedules. Among the determinants of such schedules are the amounts of food permitted at a time, the interval between feedings, and feeding in relation to the time of activity measurement. Hall and Connon (1957) predicted that different schedules, with the same amount of deprivation, would result in different amounts of activity. They found substantial differences in activity as a function of the method of handling the deprivation schedule. There is some evidence to suggest that the ratio of weight loss to predeprivation weight is more crucial than deprivation *per se*. Moskowitz (1959) was among the first to report that activity does not increase until body weight has decreased to 85-90% of normal. After this point there is almost a perfect correlation between wheel running and deprivation (prior to inanition).

Evaluation of the relationships or lack thereof between need states and activity level becomes all the more difficult when organismic variables are considered. Species differences (DeVito & Smith, 1959), strain differences (Hollifield & Parsons, 1957; McClearn, 1960), sex (Baba, 1958; Finger, 1961b), and age (Finger, 1962) have been shown to interact with either the need-activity relationship or with the measuring technique, or with both. In his rather extensive study of the effects of varying degrees and types of deprivation on different behaviors, including activity measures, Bolles (1959) comes to the conclusion that individual performance is dependent upon situational factors. There is very little communality

of performance across different situations and conditions of deprivation. Bolles (1959) states: "All of the evidence supports the hypothesis that motivated behavior is controlled chiefly by associative factors [p. 585]."

It might be well in studying activity to attempt a more refined measurement of the various types of movements the animal is making under definable need states. Perhaps such studies might reveal that what we now call general activity level is merely an artifact of inadequate measuring methods, and that activity is actually highly specific. Close analysis of the separate responses which make up the activity score may very well show that the nature of the activity being manifested is a function of the stimulus effects of different need states.

Bindra (1961) has recently attempted a formulation which is more concerned with the specificity than the generality of activity. He points out that consummatory acts must be considered in relation to the occurrence or nonoccurrence of the more pervasive and frequent spontaneous acts which comprise general activity level. He pointed out that the roles of irrelevant and relevant activity should not be ignored in analyzing the total activity matrix. Failure to do this may very well explain some of the low correlations between different measures of activity. For relevant and irrelevant response acts may display different prominence in each measure.

### CONCLUSIONS

There can be no doubt that conditions of need such as food, sex, rest, etc., influence the activity level of the organism. Most of the organic conditions are correlated with increases in general activity level. This phenomenon has been demonstrated many times. Moreover, increment in activity, up to a point, are

correlated with the degree of deficit. However, there are two crucial problems which as yet have no definite solutions. One involves the circumstances under which the relationship remains invariant, and in effect, constitutes the broader problem of definition of activity. The other one concerns the process by which need states are reflected in changes in activity.

The research findings reviewed in this paper are sufficient to indicate that any attempt to write a formula for activity output in terms of a known level of deprivation would also have to take into account the measuring device, the type of deprivation, the schedule of deprivation, the immediate environment, the experiences of the organism, the sex, age, and species (probably strain also), and the interactions between these factors. Activity, far from being a simple indicator of drive, is probably as complex a phenomenon as we have ever undertaken to measure in the laboratory. Instead of dealing with activity as a homogeneous, unitary construct, we suggest that psychologists should be devoting more efforts to the discovery of the parameters of various subclasses of activity, rather than assuming that they are measuring something which is general in nature and unitary. It is doubtful that anything is to be gained by attempting explanatory constructs that will cover the wide range of behavior included in this broad category, and the conditions under which these behaviors develop.

Probably the most commonly accepted view of need-related activity is that it represents the energizing function of drive. Accordingly activity has frequently been used to infer the presence of heightened drive state. However, in many instances, particularly in relation to food and water deprivation, an associative interpretation can also be

made. In fact, we suggest that as more is discovered about the role of learning with respect to so-called "spontaneous activity" there will be less need to refer to drive concepts. The use of constructs such as "drive" to explain certain behaviors seems to be inversely related to an understanding of the conditions under which these behaviors develop. We do not believe that the concept of drive will survive the discovery of the nature and the measurement of the antecedent organismic conditions which determine an increase in the general innervation of muscles. Learning may be the source of some of these conditions.

## REFERENCES

- ALDERSTEIN, A., & FEHRER, E. The effect of food deprivation on exploratory behavior in a complex maze. *J. comp. physiol. Psychol.*, 1955, 48, 250-253.
- AMSEL, A., & WORK, M. S. The role of learned factors in "spontaneous" activity. *J. comp. physiol. Psychol.*, 1961, 54, 527-532.
- BABA, M. An experimental study of the effect of starvation on the activity in the white rat. *Psychol. Abstr.*, 1959, 33, 744.
- BAKER, R. A. Aperiodic feeding behavior in the albino rat. *J. comp. physiol. Psychol.*, 1953, 46, 422-426.
- BARON, A., ANTONITIS, J. J., & BEALE, R. H. Effects of activity deprivation upon bar pressing. *J. comp. physiol. Psychol.*, 1961, 54, 291-293.
- BEACH, F. A. Current concepts of play in animals. *Amer. Naturalist*, 1945, 79, 523-541.
- BERLYNE, D. E. The arousal and satiation of perceptual curiosity in the rat. *J. comp. physiol. Psychol.*, 1955, 48, 238-246.
- BINDRA, D. Components of general activity and the analysis of behavior. *Psychol. Rev.*, 1961, 68, 205-215.
- BOLLES, R. C. The usefulness of the drive concept. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958. Pp. 1-33.
- BOLLES, R. C. Group and individual performance as a function of intensity and kind of deprivation. *J. comp. physiol. Psychol.*, 1959, 52, 579-585.
- BROBECK, J. R., WHEATLAND, M., & STROMINGER, J. L. Variations in regulation of energy exchange associated with estrus, di-estrus, and pseudopregnancy. *Endocrinology*, 1947, 40, 65-72.
- BRODY, E. G. The genetic basis of spontaneous activity in the albino rat. *Comp. Psychol. Monogr.*, 1942, 17, 1-24.
- CAMPBELL, B. A. Effects of water deprivation on random activity. *J. comp. physiol. Psychol.*, 1960, 53, 240-241.
- CAMPBELL, B. A., & SHEFFIELD, F. D. Relation of random activity to food deprivation. *J. comp. physiol. Psychol.*, 1953, 46, 320-322.
- CHILDERS, A. T. Hyper-activity in children having behavior disorders. *Amer. J. Orthopsychiat.*, 1935, 5, 227-243.
- DASHIELL, J. F. A quantitative demonstration of animal drive. *J. comp. Psychol.*, 1925, 5, 205-208.
- DEVITO, J. L., & SMITH, O. A. Effects of temperature and food deprivation on the random activity of *Macaca Mulatta*. *J. comp. physiol. Psychol.*, 1959, 52, 29-32.
- DUFORT, R. H., & BLICK, K. A. Adjustment of the rat to a 23-hour water deprivation schedule. *J. comp. physiol. Psychol.*, 1962, 55, 649-651.
- EAYRS, J. T. Spontaneous activity in the rat. *Brit. J. anim. Behav.*, 1954, 2, 25-30.
- ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958. Pp. 35-69.
- FEHRER, E. The effects of hunger and familiarity of locale on exploration. *J. comp. physiol. Psychol.*, 1956, 49, 549-552.
- FINGER, F. W. Hunger drive and general activity in the rat. *Amer. Psychologist*, 1949, 4, 223-224. (Abstract)
- FINGER, F. W. The effect of food deprivation and subsequent satiation upon general activity in the rat. *J. comp. physiol. Psychol.*, 1951, 44, 557-564.
- FINGER, F. W. 72 hours of food deprivation in wheels vs. photo-cage. Unpublished manuscript, University of Virginia, 1958.
- FINGER, F. W. Activity and the measurement of drive. Presidential address, Division 1, American Psychological Association, New York, 1961. (a)
- FINGER, F. W. Estrous activity as a function of measuring device. *J. comp. physiol. Psychol.*, 1961, 54, 524-526. (b)
- FINGER, F. W. Activity change under deprivation as a function of age. *J. comp. physiol. Psychol.*, 1962, 55, 100-102.
- FINGER, F. W., & REID, L. S. The effect of water deprivation and subsequent satiation upon general activity in the rat. *J. comp. physiol. Psychol.*, 1952, 45, 368-372.

- FINGER, F. W., REID, L. S., & WEASNER, N. H. The effect of reinforcement upon activity during cyclic food deprivation. *J. comp. physiol. Psychol.*, 1957, 50, 495-498.
- FINGER, F. W., REID, L. S., & WEASNER, M. H. Activity changes as a function of reinforcement under low drive. *J. comp. physiol. Psychol.*, 53, 385-387.
- FURCHGOTT, E., & ECHOLS, M. Activity and emotionality in pre- and neonatally X-irradiated rats. *J. comp. physiol. Psychol.*, 1958, 51, 541-545.
- GEIER, F. M. The measurement of tension in the rat: A contribution to method. *J. comp. Psychol.*, 1942, 34, 43-49.
- GLICKMAN, S. E., & JENSEN, G. D. The effects of hunger and thirst on Y-maze exploration. *J. comp. physiol. Psychol.*, 1961, 54, 83-85.
- HALL, J. F. Activity as a function of a restricted drinking schedule. *J. comp. physiol. Psychol.*, 1955, 48, 265-266.
- HALL, J. F. The relationship between external stimulation, food deprivation, and activity. *J. comp. physiol. Psychol.*, 1956, 49, 339-341.
- HALL, J. F. Influence of learning in activity wheel behavior. *J. genet. Psychol.*, 1958, 29, 121-125.
- HALL, J. F., & CONNOR, H. E. Activity under low motivational levels as a function of the method of manipulating the deprivation period. *J. genet. Psychol.*, 1957, 91, 137-142.
- HALL, J. F., & HANFORD, P. V. Activity as a function of a restricted feeding schedule. *J. comp. physiol. Psychol.*, 1954, 47, 362-363.
- HALL, J. F., SMITH, K., SCHNITZER, S. B., & HANFORD, P. V. Elevation of activity level in the rat following transition from ad libitum to restricted feeding. *J. comp. physiol. Psychol.*, 1953, 46, 424-433.
- HARLOW, H. F. Motivational forces underlying learning. In *Kentucky symposium: Learning theory, personality theory, and clinical research*. New York: Wiley, 1954. Pp. 36-53.
- HAYES, K. J. The effects of abnormal brain development on rat behavior. *Amer. Psychologist*, 1957, 12, 461. (Abstract)
- HILL, W. F. Activity as an autonomous drive. *J. comp. physiol. Psychol.*, 1956, 49, 15-19.
- HILL, W. F. The effect of varying periods of confinement on activity in tilt cages. *J. comp. physiol. Psychol.*, 1958, 51, 570-574.
- HILL, W. F. Effects of activity deprivation on choice of an activity incentive. *J. comp. physiol. Psychol.*, 1961, 54, 78-82.
- HOLLIFIELD, G., & PARSONS, W. Food drive and satiety in yellow mice. *Amer. J. Physiol.*, 1957, 189, 36-38.
- HUNT, J. McV., & SCHLOSBERG, H. General activity in the male white rat. *J. comp. Psychol.*, 1939, 28, 23-38.
- IRWIN, O. C. The amount and nature of activities of new-born infants under constant external stimulating conditions during the first ten days of life. *Genet. Psychol. Monogr.*, 1930, 8, 1-92.
- IRWIN, O. C. The distribution of the amount of motility in young infants between two nursing periods. *J. comp. Psychol.*, 1932, 14, 429-445.
- IRWIN, O. C. Effect of strong light on the body activity of new-borns. *J. comp. Psychol.*, 1941, 32, 233-236.
- JENKINS, J. J., & HANRATTY, J. A. Drive intensity discrimination in the albino rat. *J. comp. physiol. Psychol.*, 1949, 42, 228-232.
- KAGAN, J., & BERKUN, M. The reward value of running activity. *J. comp. physiol. Psychol.*, 1954, 47, 108.
- LINDSLEY, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 44-105.
- LUBOW, R. E., & TIGHE, T. J. A test of the discrepancy hypothesis of motivation using intensity of visual stimulation. *J. comp. physiol. Psychol.*, 1957, 50, 592-595.
- McCLEARN, G. E. The genetics of mouse behavior in novel situations. *J. comp. physiol. Psychol.*, 1959, 52, 62-67.
- McCLEARN, G. E. Strain differences in activity of mice: Influence of illumination. *J. comp. physiol. Psychol.*, 1960, 53, 142-143.
- MONTGOMERY, K. C. The effect of activity deprivation upon exploratory behavior. *J. comp. physiol. Psychol.*, 1953, 46, 438-441. (a)
- MONTGOMERY, K. C. The effect of the hunger and thirst drives upon exploratory behavior. *J. comp. physiol. Psychol.*, 1953, 46, 315-319. (b)
- MONTGOMERY, K. C. The role of the exploratory drive in learning. *J. comp. physiol. Psychol.*, 1954, 47, 60-64.
- MORGAN, C. T. Physiological mechanisms of motivations. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 1-35.
- MORGAN, C. T., & STELLAR, E. *Physiological psychology*. New York: McGraw-Hill, 1950.
- MOSKOWITZ, M. J. Running-wheel activity in the white rat as a function of combined food and water deprivation. *J. comp. physiol. Psychol.*, 1959, 52, 621-625.

- MUNN, N. L. *Handbook of psychological research on the rat*. Boston, Mass.: Houghton Mifflin, 1950.
- REED, J. D. Spontaneous activity of animals. *Psychol. Bull.*, 1947, 44, 393-412.
- REID, L. S., & FINGER, F. W. The rat's adjustment to 23-hour food-deprivation cycles. *J. comp. physiol. Psychol.*, 1955, 48, 110-113.
- REID, L. S., & FINGER, F. W. The effect of activity restriction upon adjustment to cyclic food deprivation. *J. comp. physiol. Psychol.*, 1957, 50, 491-494.
- RICHTER, C. P. A behavioristic study of the activity of the rat. *Comp. Psychol. Monogr.*, 1922, 1, 1-55.
- RICHTER, C. P. Animal behavior and internal drives. *Quart. Rev. Biol.*, 1927, 2, 307-343.
- RICHTER, C. P. Hormones and rhythms in man and animals. In G. D. Pincus (Ed.), *Recent progress in hormones research*. New York: Academic Press, 1957. Pp. 105-159.
- RICHTER, C. P., & RICE, K. K. The effect of thiamine hydrochloride on the energy value of dextrose studies in rats by the single food choice method. *Amer. J. Physiol.*, 1942, 137, 573-581.
- RICHTER, C. P., & RICE, K. K. Effects produced by vitamin D on energy, appetite and oestrous cycles of rats kept on an exclusive diet of yellow corn. *Amer. J. Physiol.*, 1943, 139, 693-699.
- ROBERTS, C. L. Stimulus and stimulus-change factors governing the free operant rate. *J. comp. physiol. Psychol.*, 1962, 55, 375-380.
- SCHULMAN, J. L., LIPKIN, N. P., CLARINDA, M., & MITCHELL, J. Studies on activity level in children. Paper read at American Psychiatric Association, Chicago, May 1961.
- SEGAL, E. F. The stability of operant level and its relation to deprivation. *J. comp. physiol. Psychol.*, 1959, 52, 713-716.
- SEWARD, J. P., & PEREBOOM, A. C. A note on the learning of "spontaneous" activity. *Amer. J. Psychol.*, 1955, 68, 139-142.
- SHEFFIELD, F. D., & CAMPBELL, B. A. The role of experience in the "spontaneous" activity of hungry rats. *J. comp. physiol. Psychol.*, 1954, 47, 97-100.
- SIEGEL, P. S. Activity level as a function of physically enforced inaction. *J. Psychol.*, 1946, 21, 285-291.
- SIEGEL, P. S., & ALEXANDER, L. E. A further observation on the effect of physically enforced inaction on the activity level of the rat. *J. genet. Psychol.*, 1948, 72, 57-62.
- SIEGEL, P. S. & STEINBERG, M. Activity level as a function of hunger. *J. comp. physiol. Psychol.*, 1949, 42, 413-416.
- SPECTOR, H., & YOUNG, D. R. Physical performance capacity and nutriture: Evaluation of rations by animal experimentation. *Amer. J. clin. Nutr.*, 1957, 5, 129-140.
- STERN, J. A. The effect of frontal cortical lesions on activity wheel and open-field behavior. *J. genet. Psychol.*, 1957, 90, 203-212.
- STEVENSON, J. A., & RIXON, R. H. Environmental temperature and deprivation of food and water on the activity of rats. *Yale J. Biol.*, 1957, 29, 575-584.
- STONE, C. P., & BARKER, R. G. Spontaneous activity, direct and indirect measures of sexual drive in adult male rats. *Proc. Soc. Exp. Biol. N. Y.*, 1934, 32, 195-199.
- STRONG, P. N., JR. Activity in the white rat as a function of apparatus and hunger. *J. comp. physiol. Psychol.*, 1957, 50, 596-600.
- TAINTER, M. L. The effects of certain analeptic drugs on spontaneous running activity of the white rat. *J. comp. Psychol.*, 1943, 36, 143-155.
- TEGHTSONIAN, R., & CAMPBELL, B. A. Random activity of the rat during food deprivation as a function of environmental conditions. *J. comp. physiol. Psychol.*, 1960, 53, 242-244.
- TEITELBAUM, P. Random and food-directed activity in hyperphagic and normal rats. *J. comp. physiol. Psychol.*, 1957, 50, 486-490.
- TREICHLER, F. R., & HALL, J. F. The relationship between deprivation weight loss and several measures of activity. *J. comp. physiol. Psychol.*, 1962, 55, 346-349.
- WALD, G., & JACKSON, S. Activity and nutritional deprivation. *Proc. Nat. Acad. Sci., Washington*, 1944, 30, 255-263.
- WANG, G. H. The relation between "spontaneous" activity and oestrous cycle in the white rat. *Comp. Psychol. Monogr.*, 1923, 2, 1-27.
- WEASNER, M. H., FINGER, F. W., & REID, L. S. Activity changes under food deprivation as a function of recording device. *J. comp. physiol. Psychol.*, 1960, 53, 470-474.
- WEBB, W. B. A study of random bar pressing responses as a function of drive states. *Amer. Psychologist*, 1948, 3, 333. (Abstract)
- WOLFENBERGER, W. P., MILLER, M. B., FOSHEE, J. G., & CROMWELL, R. L. Rorschach correlates of activity level in high school children. *J. consult. Psychol.*, 1962, 26, 269-272.

(Received May 1, 1963)

# PHONETIC SYMBOLISM RE-EXAMINED

JONATHAN H. WEISS

*Children's Asthma Research Institute and Hospital, Denver, Colorado*

Taylor (1963) pointed out (a) results of studies of universal phonetic symbolism in natural languages were contradictory (b) Taylor and Taylor (1962) showed that speakers of different languages tend to assign different meanings to the same sounds. She argued, therefore, that current theories of phonetic symbolism, which imply its universality, must be revised, and offered a theory based upon the occurrence of certain sounds more predominantly in some meaning categories than in others. It is pointed out, however, that Taylor overlooked important methodological factors in "contradictory" studies and did not take into account findings contrary to those of Taylor and Taylor. In addition, the "language habits" theory is examined and various untested assumptions and other difficulties are pointed out.

In a recent review of the literature, Taylor (1963) argued for a basic revision of phonetic symbolism theories because they imply that the phenomenon is universal and experimental evidence has failed to bear this out. Specifically, Taylor argued that theories which postulate that the sound-meaning relations to which phonetic symbolism refers are based upon such things as learned associations between sounds and object types in the world (Brown & Nuttall, 1959), the natural properties of sounds (Bentley & Varon, 1933; Miron, 1961) or "physical scales" of sound production (Newman, 1933; Sapir, 1929), should predict that all people, regardless of their native language, attach the same meanings to sounds. However, the results of studies designed to test this, indirectly or directly, have, according to Taylor, been either inconclusive or directly in opposition to the hypothesis. For example, in one type of study of universal phonetic symbolism in natural languages, subjects may be asked to guess the meanings of foreign words by matching the corresponding members of word pairs in the subjects' native language and in some noncognate foreign language. Where no structural cues can be demonstrated, correct guessing is assumed to be based upon

the subjects' response to the meaning implications of sounds of the foreign words. In two out of six such studies reviewed by Taylor (Brackbill & Little, 1957; Maltzman, Morrisett, & Brooks, 1956), essentially negative results led the investigators to conclude that phonetic symbolism was no symbolism at all. In a second type of study, designed to test the universal-phonetic-symbolism hypothesis directly, Taylor and Taylor (1962) asked subjects in four different language communities to rate 144 nonsense syllables on four dimensions of meaning. Results seemed to indicate that the same sounds had different meanings for people who speak different languages.

On these grounds, then, Taylor (1963) postulates that phonetic symbolism, rather than being universal, must be culturally delimited and therefore based upon "some factors that are different from one language (community) to another unrelated language (community) [p. 206]." She postulates that the likely factor is language habits. The specific language habit which Taylor proposes derives from what she believes is an established observation, namely, that words sharing a common connotation tend to have the same initial consonant. This consonant, accord-

ing to Taylor, through association with the connotation of the words in which it occurs takes on that connotation. Thus, English-speaking people learn that *g* is a "big sound" because in English *g* is more often the initial consonant of words connoting "bigness" than of words connoting "smallness." Similarly, *t* connotes "smallness" because it is more often the initial consonant of "small words."

Having summarized Taylor's position, we may proceed first to examine the grounds on which Taylor's objection to current theories is based and then undertake to evaluate Taylor's "language habits" theory in its own right.

#### THE GROUNDS FOR TAYLOR'S OBJECTION

First, the negative results in the word-matching studies of Brackbill and Little (1957) and Maltzman et al. (1956) may be considered. In the opinion of the present author these results are taken uncritically by Taylor as evidence against universal phonetic symbolism. For, as Brown and Nuttall (1959) and Weiss (1963) have pointed out, both Brackbill and Little and Maltzman et al. introduced important procedural variables which may have accounted for their results. Thus, Brackbill and Little asked subjects to tell if a single stimulus and a single response word were the "same" or "different" in meaning, and thereby effectively reduced to one fourth the amount of information available to the subject in this situation as opposed to one in which word pairs were used. Maltzman et al., on the other hand, used foreign words in both the stimulus and response categories, and Weiss (1963) has shown that it is necessary that the words in at least one of the categories be meaningful to the subject. Nowhere does Taylor consider these points. On the contrary, her observa-

tion that "the more recently the experiments were done the less the degree of positive results," and her suggestion that "perhaps more recent experiments have tended to eliminate sources of spurious positive findings [1963, p. 201]" may betray a bias. Thus, the possibility of "spurious negative findings," if findings can ever be considered spurious at all, is not considered in Taylor's evaluation of the evidence.

Second, as far as the Taylor and Taylor (1962) study is concerned, the findings do indeed present a difficulty for the concept of universal phonetic symbolism. Thus, in word-matching studies the interpretation of correct guessing is that while different languages may use different sounds to convey the same meaning, the symbolic values of the sounds in a foreign word are universally shared and therefore make it possible for nonspeakers of the foreign language to guess the meaning of that word. Taylor and Taylor's findings, however, make this seem an unlikely possibility. Is there an explanation, then, which can reconcile both findings without incorporating the notion of universal phonetic symbolism? One attempt has been made by Brackbill and Little (1957) who argue that the choices in word-matching experiments are based upon subjects' perception of structural similarities between the words guessed to have the same meaning. If this were true, it would indeed argue against the appropriateness of considering the subjects' responses in that paradigm as being based upon phonetic symbolism. However, it must be recognized that while the argument may explain significant agreement, it does not explain correct agreement. In other words, it does not explain how it happens that different languages make use of similarly constructed words to convey the same meaning. In fact, it is the opinion

of the present author that structural similarity in historically unrelated words of the same meaning is in itself a potent argument for some process of symbolism. Therefore, unless one were willing to grant that the positive findings of word-matching studies are indeed spurious, an alternative theory remains to be advanced which could accommodate both positive findings in word-matching studies and findings such as those of Taylor and Taylor (1962). It should be noted, however, that there are at least two very recent findings which are in some measure contrary to those of Taylor and Taylor. Davis (1961) found that African children who had never been exposed to English tended to assign the nonsense words *TAKETE* and *ULOOMU* to angular and rounded pictures in much the same way as did British children, and Miron (1961) found a high degree of similarity in ratings of nonsense syllables across two language groups (English and Japanese). Although it is true that Miron's (1961) Japanese subjects had varying degrees of familiarity with English, he concluded that,

the materials had expressive symbolic value accruing to their inherent phonetic content and not to any meanings via real-word associates. These affective meanings were found to bear consistent, lawful relations to the phonetic properties of the sounds. The fact that these meaningful differentiations and their relations to phonetic properties proves to be highly similar across two contrasting linguistic groups suggests that the laws governing phonetic symbolism may have a universal character [p. 630].

In light of these findings one is led to speculate that since the theory of phonetic symbolism does not specify that a single meaning becomes associated with a single sound, but rather that many meanings may become hierarchically associated with a sound, and vice versa, the hierarchies established by different cultural groups may differ, pos-

sibly in part as a result of exposure to such language habits as Taylor (1963) describes. Thus, in the Taylor and Taylor (1962) study the meaning attached by each language group to a given sound may simply have been the dominant meaning in its hierarchy of associations to that sound. While this explanation is not entirely satisfactory in light of the fact that Taylor and Taylor (1962) found no appreciable correlations across language groups between the ratings of many sounds on the same dimension, it seems to offer at least the core of a reconciliation. In any case, it is quite clear that there is a need for further exploration.

#### THE LANGUAGE-HABITS THEORY

Since it is possible that alternative explanations may exist at this time for the data of phonetic-symbolism studies, Taylor's language-habits theory of phonetic symbolism deserves to be evaluated in its own right. While at first glance the theory seems to be a plausible explanation for the Taylor and Taylor (1962) findings, it has a number of difficulties which need to be considered. First of all, the theory as stated begs the question of how it comes about in the first place that unrelated words in a given language which share a connotative meaning tend also to share structural components. It seems that rather than phonetic symbolism being the result of such linguistic phenomena as are described by Taylor, just the reverse may be true. That is, because of the symbolic values acquired by sounds through extralinguistic associations, different words which represent a meaning category, or refer to it connotatively, draw upon the same sounds because of the association of those sounds to that category. As it stands, Taylor's theory cannot explain the observation of such people as Jespersen (1922) that some

word types, for example, words of action, words related to sounds, and words of imitation, seem to have more phonetic symbolism. The hypothesis of extra-linguistic associations, on the other hand, could account for it quite well.

A second difficulty of Taylor's theory is that while it proposes that phonetic symbolism is based upon an association between word meanings and the initial consonants of words, it offers no reason why initial consonants should be the ones to carry the burden of the proposed associations. Thus, Horowitz (1955) has pointed out that in most Indo-European languages it is the final syllable or suffix which usually carries the adjectival component of a word. Why then do speakers not learn to associate the final sounds of a word with its connotations? Furthermore, Taylor offers no evidence that initial consonants of a word in a meaning category are in fact more frequently similar than are any other components of those words.

Third, the evidence offered by Taylor that a larger number of words of a given connotation have the same initial consonant than do words of the opposite connotation, is open to the very criticism of selectivity that Taylor rightly levels against such people as Paget (1930) who showed that some words are constructed in a manner consistent with what their theories would predict. The words which do not fit the model are never mentioned. Thus, to take the list of words offered by Taylor, it is true that there are many more big words beginning with the G sound or K sound (GARGANTUA, GLARING, GOODLY, GRAVE, GREAT, GROSS, CAPACITY, CARGO, COLOSSUS, COMPREHENSIVE, CONSIDERABLE, CONSUMMATE, CORPORATION, CORPULENCE, CRASS, and QUANTITY) than there are small words having those initial consonants (GLEAM, GLOBULE, GRAIN, GRUB, CLOSE, CONTRACT, CORPUSCLE,

CRUMB, and CRUMBLE) and Taylor's hypothesis would seem to be borne out. However, one can easily think of a sufficient number of small words beginning with G and K (e.g., CRIB, CRAWL, CAN, GRAM, GUN, GLASSES, and so on) which, if included in the list, would cause the numerical difference to vanish, if not, in fact, to be reversed. The selection of words as it now stands, is entirely arbitrary.

On the whole, then, the promulgation of a new theory of phonetic symbolism on the grounds put forward by Taylor should at least await further confirmation of those grounds. However, Taylor's work serves well to indicate the necessity for further cross-cultural work on universal phonetic symbolism. In fact, more systematic study of the parameters of phonetic symbolism as a response capacity is similarly indicated. Unfortunately, at this time there are too many cases being made in defense of too few facts.

## REFERENCES

- BENTLEY, M., & VARON, EDITH. An accessory study of "phonetic symbolism." *Amer. J. Psychol.*, 1933, 45, 76-86.
- BRACKBILL, YVONNE, & LITTLE, K. B. Factors determining the guessing of meaning of foreign words. *J. abnorm. soc. Psychol.*, 1957, 54, 313-318.
- BROWN, R., & NUTTALL, R. Methods in phonetic symbolism experiments. *J. abnorm. soc. Psychol.*, 1959, 59, 441-445.
- DAVIS, R. The fitness of names to drawings: A cross-cultural study in Tanganyika. *Brit. J. Psychol.*, 1961, 52, 259-268.
- HOROWITZ, A. E. The effect of variation in linguistic structure on the learning of miniature linguistic systems. Unpublished doctoral dissertation, Harvard University, 1955.
- JESPERSEN, O. *Language*. New York: Holt, 1922.
- MALTZMAN, I., MORRISSETT, L., & BROOKS, L. O. An investigation of phonetic symbolism. *J. abnorm. soc. Psychol.*, 1956, 53, 249-251.
- MIRON, M. S. A cross-linguistic investigation of phonetic symbolism. *J. abnorm. soc. Psychol.*, 1961, 62, 623-630.

- NEWMAN, S. S. Further experiments in phonetic symbolism. *Amer. J. Psychol.*, 1933, 45, 53-75.
- PAGET, R. *Human speech*. New York: Harcourt, Brace, 1930.
- SAPIR, E. A study in phonetic symbolism. *J. exp. Psychol.*, 1929, 2, 225-239.
- TAYLOR, INSUP K. Phonetic symbolism re-examined. *Psychol. Bull.*, 1963, 60, 200-209.
- TAYLOR, INSUP K., & TAYLOR, M. M. Phonetic symbolism in four unrelated languages. *Canad. J. Psychol.*, 1962, 16, 344-356.
- WEISS, J. H. Role of "meaningfulness" versus meaning dimensions in guessing the meanings of foreign words. *J. abnorm. soc. Psychol.*, 1963, 66, 541-546.

(Received May 15, 1963)

# LEARNING AND PSYCHOPATHOLOGY

FREDERICK W. HUFF<sup>1</sup>

University of Georgia

Experiments concerning the learning of abnormal human Ss are reviewed. The review includes discussion of how persons with particular types of psychopathology learn different kinds of tasks as well as the reasons suggested for the learning phenomena found.

This paper is an attempt to bring together most of the learning experiments using abnormal human subjects. It has two purposes: to provide a review of the literature and to bring to light seemingly valid generalizations.

## LEARNING OF PSYCHOTICS

### *Schizophrenics*

*Paranoids.* No studies were found which investigated exclusively the learning of paranoid schizophrenics. Huston (1932), Huston and Shakow (1936), and Shakow (1932) have demonstrated that of the schizophrenic subtypes, paranoids are more like normals than the others, although they are somewhat inferior to normals. The tasks employed were the prod meter (Shakow, 1932), speed of tapping (Huston & Shakow, 1936), and the pursuitmeter (Huston, 1932).

Several studies (O'Connor & Rawnsley, 1959; Pfaffman & Schlosberg, 1936) utilizing classical conditioning have failed to duplicate these results. Pfaffman and Schlosberg (1936), in conditioning the knee jerk, found no differences between the subcategories of schizophrenia. Using eyelid conditioning, O'Connor and Rawnsley (1959) also demonstrated that paranoids did not differ from other schizophrenics. However, Pfaffman and Schlosberg (1936) showed that paranoids conditioned more poorly than normals,

whereas O'Connor and Rawnsley (1959) found no difference. These conflicting results could be due to subject variables other than paranoia. To add to the confusion, O'Connor (1957) and O'Connor and Rawnsley (1959) demonstrated on other tasks (crossing out odd and underlining even numbers and verbal conditioning, respectively) that paranoid schizophrenics were slow learners when compared with nonparanoid schizophrenics.

That paranoids may perform uniquely on learning tasks is supported by the fact that some investigators are careful to exclude them when studying the learning of other schizophrenics (Tizard & Venables, 1957; Venables & Tizard, 1956b, 1956c). Further, it has been shown that on a complex reaction-time task nonparanoid schizophrenics improved when an extraneous stimulus (white noise) was introduced, while paranoids did not (Tizard & Venables, 1957).

*Hebephrenic and Mixed Types.* In the Huston and Shakow studies mentioned above, hebephrenic and mixed schizophrenics were found to be inferior in performance to paranoids. However, the relationship did not hold for classical conditioning (Pfaffman & Schlosberg, 1936).

*Catatronics.* Catatonic schizophrenics are the poorest learners among all the schizophrenic subtypes on the tasks employed in the Huston and Shakow studies, but their conditionability was the same as the other subtypes (O'Connor

<sup>1</sup> Acknowledgments are due Robert E. Taylor for his able editorial assistance.

nor & Rawnsley, 1959; Pfaffman & Schlosberg, 1936). However, Shipley (1934), using GSR conditioning, indicated that catatonics conditioned better than normals.

*Unspecified.* This section deals with studies in which the type of schizophrenia is unspecified. It is likely that the subjects were a mixture of all types. Nevertheless, it would not be correct to think that the subjects were representative of the total schizophrenic population. They were chosen on the basis of availability and capacity to participate. Also, because of the nonspecificity of the variables defining schizophrenia it is difficult to compare studies. Since there may be differences in the performance of the schizophrenic subtypes on learning tasks, the present studies are thus confounded in not taking this into account. However, since with one exception (O'Connor & Rawnsley, 1959), the available data indicate that all schizophrenics are poorer on learning tasks than normals, they perhaps err only in the magnitude of the learning deficit found.

Kent (1911), using many verbal-learning tasks, was one of the first to demonstrate that schizophrenics could learn. Boring (1913) confirmed this, but found that schizophrenics were inferior when compared with normals. This deficiency also occurred in paired-associate learning (Atkinson, 1958; Hall & Crookes, 1951; Hull, 1917; Robinson, 1958), but retention did not seem to be impaired (Hull, 1917).

There has been some exploratory work on the verbal conditioning of schizophrenics. There are no comparisons of schizophrenics with normals (Cohen & Cohen, 1960; Salzinger & Pisoni, 1958). Salzinger and Pisoni (1958) were able to condition "I" and "we" responses in schizophrenics using agreement as reinforcement. Cohen and Cohen (1960) tried to condition first-

and second-person pronouns and were unable to do so. However, they used a different reinforcer (the word "good") and failed to control variables known to affect the learning of schizophrenics; for example, cooperation, severity of illness, etc. Thus, it is difficult to compare the two studies. More recently, Salzinger and Pisoni (1961) have duplicated their earlier findings.

Reaction-time studies have been included here because of their obvious relationship to motor learning, and because they show changes with practice, and thus can be considered as affected by learning. (The reader is referred to King, 1954, for a general treatment of the psychomotor aspects of mental disease.) Many studies (Huston & Shakow, 1936; Huston, Shakow, & Riggs, 1937; Rodnick & Shakow, 1940; Venables & Tizard, 1956c) have reliably shown that schizophrenics have a slower reaction time than normals, and that with practice little improvement occurs. This holds for a wide variety of reaction-time tasks, ranging from simple tapping, to manipulating complex switches to cues provided by different colored lights. Generally, as intensity of cue stimulus increases so does speed of reaction time; however, Venables and Tizard (1956b) found that with chronic, long-term schizophrenics the reverse was true where the cue stimulus was a light. However, on further investigation this "paradoxical" effect could not be duplicated (Venables & Tizard, 1956a).

On the pursuitmeter, schizophrenics started at a lower initial level than normals and improved with increasingly smaller increments, the curves appearing more negatively accelerated than with normals (Huston, 1932; Huston & Shakow, 1948). Peters (1956), using circular pencil mazes, found schizophrenics poorer in both time and errors.

The operant conditioning of gross motor behavior in schizophrenics is fairly

recent. They had a slower rate of acquisition than normals and under conditions of reinforcement they had long periods of nonresponding. In a few subjects the conditioned response was found to be extremely resistant to extinction, even after many hours of nonreinforcement (Lindsley, 1956). Bullock (1960) also indicates that schizophrenics are inferior to normals on operant conditioning.

The empirical findings using the classical conditioning paradigm are contradictory as to the conditionability of the schizophrenic, especially among the subcategories (O'Connor & Rawnsley, 1959; Pfaffman & Schlosberg, 1936). Howe (1958) found conditionability of psychogalvanic response (PGR) to be no different in schizophrenics and normals, while Peters and Muphree (1954) collected evidence to the contrary. In a review of the relevant Russian literature, Lynn (1963) cites findings which support Howe's conclusions.

### *Depressives*

Only a few studies (Bregelmann, 1957; Venables, 1959; Venables & Tizard, 1956c) have included "psychotic depressives," and in these the criteria for selection of subjects varies. Two on reaction time (Venables & Tizard, 1956c; Venables, 1959) indicate that depressives start at a lower level than normals (even lower than schizophrenics), but that they finally achieve a level commensurate with normal subjects. In an investigation where the task was to reproduce visual patterns after a short exposure, error scores distributed bimodally, perhaps suggesting two types of depressives (reactive and endogenous). Depressives were inferior to normals on this task (Bregelmann, 1957).

### *Manic-Depressives*

Manic-depressives, included in two of the Huston and Shakow studies (1948,

1949), had no impairment in learning. On GSR conditioning they conditioned somewhat better than normals (Shipley, 1934). Similar findings are reported by Pfaffman and Schlosberg (1936) in conditioning the knee jerk.

### *Unspecified*

A number of workers (Mailloux & Newburger, 1941; Mednick & Lindsley, 1958; Spence & Taylor, 1953; Taylor & Spence, 1954) used psychotic patients without any further specification as to subject population. Since there appears to be differences in the learning of different types of psychotics, this kind of study almost evades interpretation.

On a discrimination reaction-time task, psychotics were able to learn to discriminate between colored lights (Mailloux & Newburger, 1941), but because of the heterogeneous population and lack of control group, it is uncertain just what this means. On two occasions of eyelid conditioning it was demonstrated that a psychotic group could be conditioned better than normals (Spence & Taylor, 1953; Taylor & Spence, 1954). The reason for these findings is unknown. It might be that the psychotic group included many manic-depressive subjects who have been shown to condition better than normal subjects (Shipley, 1934; Pfaffman & Schlosberg, 1936), or it might be that schizophrenics who presumably make up the bulk of any psychotic group do condition better than normals, but this seems unlikely in face of the evidence already cited. Mednick and Lindsley (1958) report much response variability in operant motor conditioning with psychotics.

### *Variables Affecting the Learning of Psychotics*

The studies to be reviewed here were chosen because they explored, experimentally, factors which might explain the learning phenomena presented above.

Strictly theoretical papers were omitted. These studies were primarily with schizophrenic subjects, and thus, most of the variables apply to schizophrenic learning. They might also account for the learning of other abnormal subjects, but the empirical evidence is lacking. The variables mentioned here do not in any way exhaust the theoretical possibilities.

*Interference.* Three investigations (Bleke, 1954; Lang & Luoto, 1962; Mednick & Lindsley, 1958) have demonstrated that intrusion of irrelevant (wrong) responses is greater in the learning of psychotics and schizophrenics than normals. One of these (Bleke, 1954) has gone further and shown that reminiscence is greater for schizophrenics than normals, thus giving indirect evidence of increased interference. He had schizophrenics learn a temporal maze consisting of switches where the discriminative stimuli were words, and found that reminiscence was greater for schizophrenics with the poorest premorbid adjustment, but not for the others. A normal control group showed less reminiscence than those with poor premorbid adjustment, but not the remaining schizophrenics. This suggests another variable, severity of illness, which might be a factor in the learning of schizophrenics; the experimental evidence on this point will be reviewed later. Psychotics in general made more irrelevant and nonpurposive responses than did normals on operant motor conditioning (Mednick & Lindsley, 1958). Lang and Luoto (1962) tested the verbal mediation hypothesis using the paired-associate method. Schizophrenics showed more intrusions of previously learned words than did normals. It appears that remote associations are greater in schizophrenics, although this hypothesis has not been tested directly. It seems that schizophrenics, and psychotics in general, are poorer in learning

tasks than normals because of heightened intrusion of irrelevant (wrong) responses.

*Overgeneralization.* Several workers (e.g., Cameron & Margaret, 1951) have suggested that schizophrenics overgeneralize. This does not seem to be true when schizophrenics are taken as a whole; however, it might be true for specific subtypes. Bender and Schilder (1930) found overgeneralization along a pitch dimension in a group of catatonic schizophrenics when compared with normals, while Mednick (1955), using a visual stimulus, obtained no differences when unspecified schizophrenics were used. Garmezy (1952) did not find any differences in visual discrimination of normals and schizophrenics, thus suggesting no difference in stimulus generalization.

*Responsiveness.* Using several different measures, there is evidence that schizophrenics are physiologically less responsive than normals. The other psychotic groups have not been studied from this point of view. However in one study (Huston & Shakow, 1936) there was no difference in the speed of tapping of manic-depressives and normals. Speed of tapping and reaction time are both slower in schizophrenics than in normals (Huston & Shakow, 1936; Huston, Shakow, & Riggs, 1937; Rodnick & Shakow, 1940; Venables & Tizard, 1956c). Speed of tapping also varied according to type of schizophrenia with the results distributing themselves as they did with reaction time; that is, paranoids were the fastest, catatonics slowest, with mixed and hebephrenic types intermediate (Huston & Shakow, 1936). Also, in support of lowered responsiveness or excitability, Tizard and Venables (1957) found that a simple increase in the level of stimulation (illumination of room) caused a significant increase in the reaction time of non-

paranoid schizophrenics. This effect was not obtained with paranoid schizophrenics. Peters and Murphree (1954) report the unconditioned PGR was less for schizophrenics than normals. Winokur, Guze, Stewart, Pfeiffer, Stern, and Hornung (1959) have even suggested that perhaps all abnormal subjects could be differentiated in terms of unconditioned PGR, and they present evidence to support this. They obtained a positive relationship between the number of unconditioned PGR responses given to an alerting stimulus (tone) and rate of classical conditioning. In a study of general physiological responsiveness, the schizophrenic was not different from normal controls (Malmo, Shagass, & Smith, 1951).

*Reactive Inhibition.* Many workers (Huston & Shakow, 1936; Huston, Shakow, & Riggs, 1937; O'Connor, 1957; Venables & Tizard, 1956c) have stated that psychotics show increasing work decrement and overall low level of responding because of heightened susceptibility to reactive inhibition. This hypothesis has been tested in two ways: studying disinhibition by introducing extraneous stimuli into the learning process and studying reminiscence. If reactive inhibition were greater in schizophrenics than in normals, then disinhibition would be expected to be greater. This has been obtained using extraneous stimuli (white noise) as a disinhibitor and reaction time as the dependent variable (Lang, 1959; Pascal & Swensen, 1952; Tizard & Venables, 1957). However, other studies indicate that this heightened disinhibition can be explained in terms other than reactive inhibition. Karras (1962), duplicating Lang's study, demonstrated that the introduction of extraneous stimuli served as added reinforcement; that is, the subject was motivated to escape the white noise and thus reaction time was de-

creased. He maintained that in Lang's procedural setup the white noise was terminated shortly after a response was given, thus acting as reinforcement. In another study (Tizard & Venables, 1957), a facilitating effect was obtained by just increasing background stimulation. Thus, the facilitating effect of the introduction of extraneous stimuli might be the result of either the escape nature of the task, or of increased arousal, or both. Further, if reactive inhibition were greater for schizophrenics, then reminiscence would be expected to be greater. Several workers (O'Connor 1957; Venables, 1959) indicate that this is not the case when simple motor (O'Connor, 1957) and reaction-time tasks (Venables, 1959) are used.

*Reinforcement.* The nature of the reinforcing stimulus situation is an important variable in all learning studies, perhaps especially so in studies using abnormal subjects. Although not examined directly, there is evidence to indicate that what constitutes reinforcement often varies for abnormal and normal subjects. In a study on operant motor conditioning in psychotics (Lindsley, 1956), normal subjects worked well for certain rewards whereas psychotic subjects did not.

Schizophrenics might learn better under escape conditions than reward (Cavanagh, 1958; Cohen, 1956; Karras, 1962). Cohen (1956) taught schizophrenics two temporal mazes by shocking the incorrect responses. Shock improved learning of the second maze only, and had no effect on the learning of normals. It is not clear if this represents real improvement, or only recovery from the initial disruptive effects of shock. If schizophrenics are overly responsive to shock they might show a learning deficit on the first maze, but improved learning on the second due to adaptation. Schizophrenics produced more nor-

mal concepts on a concept formation test when a correct response, that is, a more "normal" concept, terminated white noise than when it did not (Cavanagh, 1958). Might a reduction of stimulation be more reinforcing for schizophrenics than normals?

In a well-controlled experiment using schizophrenics, Karras (1962) found that avoidance of a loud tone was more reinforcing than nonavoidance in reducing reaction time, also, the more intense the tone the shorter the latency. Even when the tone could not be avoided, reaction time was somewhat reduced. Karras concluded that improvement was a function of both reinforcement and arousal, and that if either one were low or lacking, learning would be impaired. On the surface it appears that these findings are contradicted by studies in which punishment was used. Several workers (Aisenberg, 1957; Garmezy, 1952) have reported that punishment hampers learning, but others have found facilitation (Atkinson, 1958; Robinson, 1958). The reason for these inconsistent findings is not clear. Punishment and task differed as did subject sample, so the relevant variables have not been isolated. Superficially, it appears that when the schizophrenic can lower the general level of stimulation by learning, learning is facilitated, but that when this does not occur learning is hampered.

*Meaning.* It is axiomatic that the meaning ascribed to stimuli by psychotics differs from that of normal persons. As a matter of fact, abnormality is often defined in terms of the unusualness with which a standard stimulus is interpreted. It seems reasonable then that meaning would differentially affect learning. Davis and Harrington (1957) have shown that human stimuli disrupt problem solving in schizophrenics. In another study (Gladis & Wischner, 1962), it was more difficult for schizophrenics to learn

"aversive" words than neutral ones in a paired-associate task. One might postulate that schizophrenics have difficulty learning because a great many stimuli become aversive. This could account for the fact that schizophrenics learn best under avoidance conditions. Here the avoidance is rewarding because many stimuli are aversive.

*Anxiety.* Many studies show that clinical and manifest (test) anxiety variously affect the course of learning many types of tasks (Diethelm & Jones, 1947; Gilberstadt & Davenport, 1960; Hunt, 1959; Korchin & Levine, 1957; Welch & Kubis, 1947b). Although no investigation was found in which anxiety was systematically varied in psychotic patients, it seems reasonable to assume that anxiety has the same effect on their learning as it does on that of neurotics. Using mixed neurotics and psychotics, anxiety seems to increase conditionability (Gilberstadt & Davenport, 1960; Welch & Kubis, 1947b) but to impair complex learning (Diethelm & Jones, 1947; Korchin & Levine, 1957), except when noxious cues are present, and then learning is facilitated (Rosenbaum, 1956). Thus, anxiety has different effects depending upon: the type of learning task; complexity of task; and nature of reinforcing conditions, for example, avoidance, punishment, reward, etc. These studies used heterogeneous patient groups and did separate neurotics from psychotics, etc.

*Severity of Illness.* There are many indications that severity of illness is inversely related to learning performance; that is, the "deeper" the psychosis and the more chronic, the poorer the learning (Hall & Crookes, 1951; Mednick & Lindsley, 1958; Peters, 1956; Smith, 1959). In one study (King, Merrell, Lovinger, & Denny, 1957) this did not hold, but here only acute schizophrenics were used, and since acute schizophren-

ics have been shown to be alike in learning performance (Long, 1960) the presumption is that they are all equally ill.

*Motivation.* The poor learning of schizophrenics might be due in part to their lack of cooperation or proper motivation (Cohen, 1956; Huston & Shakow, 1946; Kent, 1911; Tizard & Venables, 1957). Psychotics, especially schizophrenics, have all been thought to be deficient in motivation, cooperation, and attention. Unfortunately these concepts have not been differentiated in operational terms; thus, it is not possible at this point to decide between them.

#### LEARNING OF NEUROTICS

Few studies have been conducted concerning the learning of neurotics. This is probably due in part to the difficulties involved in obtaining a neurotic group. Since they are not hospitalized it is difficult to observe them over extended periods of time. Also, on theoretical grounds there is less reason to expect learning deficiencies in neurotics, and therefore interest has been somewhat less than it has been in psychotics.

##### *Anxiety Neurotics*

It appears that anxiety neurotics are inferior to normals in verbal learning, both rote serial (Malmo & Amsel, 1948) and paired-associate (Hall & Crookes, 1952). Because the subject population is not clearly defined it is difficult to compare the results of different experiments on the conditioning of anxiety neurotics. Taylor and Spence (1954) did not find any differences in the conditioning of anxiety neurotics, other neurotics, and normals. Yet in another experiment (Franks, 1956), anxiety neurotics conditioned better than hysterics who have been shown to be inferior to normals (Hilberstam, 1961). Hilberstam (1961) also reported that introverted neurotics condition better

than normals. Are these anxiety neurotics? Two variables are probably confused here, psychiatric nosology and introversion-extraversion (cf. Eysenck, 1962). Anxiety neurotics conditioned better than normals when magnitude of the conditioned response during extinction was used as a measure of conditioning (Howe, 1958).

##### *Hysterics*

Hysterics condition differently from normals and other neurotics (Franks, 1956; Hilberstam, 1961). Using eyeblink and PGR, hysteric subjects conditioned poorer than subjects with anxiety states (Franks, 1956). Another PGR study found hysteric subjects inferior to normal subjects (Hilberstam, 1961).

##### *Depressives and Obsessive-Compulsives*

Hall and Crookes (1952) reported that depressives were consistently inferior to normals on paired-associate learning as were obsessive-compulsives. They did not investigate differences between depressives and other neurotics.

##### *Unspecified*

The studies reported here used neurotic subjects taken as a group without regard to type of neurosis. Several (Spence & Taylor, 1953; Taylor & Spence, 1954) have failed to find any differences in the conditioning of neurotics and normals. One was found of verbal learning and here "emotionally disturbed" boys were used (Davids & Oliver, 1960). Davids and Oliver used a paired-associate task and disturbed boys were inferior to normal controls. They also used a finger maze, but in contrast, the boys were superior to normals on this task. Brengelmann (1957) obtained similar results using adult neurotics on a figure-reproduction task. Neurotics appear inferior to normals in

operant conditioning of gross motor behavior (Bullock, 1960). Cohen and Cohen (1960) have demonstrated that neurotics can be verbally conditioned, but because no control group was used it is unknown if they differ from normals.

### Character Disorders

Lykken (1957) studied the conditioning (GSR) of psychopaths, and found that primary psychopaths (chronic with no apparent precipitating cause) were significantly poorer than neurotic psychopaths and normals.

### Variables Affecting the Learning of Neurotics

The reasons advanced to explain the learning of neurotic subjects are similar to those used to explain the learning of psychotic subjects. Among the more common ones explored experimentally are: reactive inhibition (Malmo & Wallerstein, 1955), connotative meaning of words in a paired-associate task (Davids & Oliver, 1960), interference from previously learned material (Davids & Oliver, 1960; Malmo & Amsel, 1948), physiological reactivity (Winokur et al., 1959), and anxiety (Hunt, 1959).

### REFERENCES

- AISENBERG, R. B. The cumulative differential effects of reward and punishment on the performance of schizophrenics and normal subjects. Unpublished doctoral dissertation, Boston University, 1957.
- ATKINSON, R. L. Paired-associate learning by schizophrenics and normal subjects under conditions of verbal reward and verbal punishment. Unpublished doctoral dissertation, Indiana University, 1958.
- BENDER, L., & SCHILDER, P. Unconditioned reactions to pain in schizophrenics. *Amer. J. Psychiat.*, 1930, 10, 365-384.
- BLEKE, R. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *J. Pers.*, 1954, 23, 479-498.
- BORING, E. G. Learning in dementia praecox. *Psychol. Monogr.*, 1913, 15(2, Whole No. 63).
- BRENGELMANN, J. C. Learning in neurotics and psychotics. *Acta psychol., Amsterdam*, 1957, 13, 371-388.
- BULLOCK, D. H. Performances of psychiatric patients in a brief operant discrimination test. *Psychol. Rec.*, 1960, 10, 83-93.
- CAMERON, N., & MARGARET, A. *Behavior pathology*. Boston, Mass.: Houghton Mifflin, 1951.
- CAVANAGH, D. Improvement in the performance of schizophrenics on concept formation tasks as a function of motivational change. *J. abnorm. soc. Psychol.*, 1958, 57, 8-12.
- COHEN, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 52, 186-190.
- COHEN, E., & COHEN, B. D. Verbal reinforcement in schizophrenia. *J. abnorm. soc. Psychol.*, 1960, 60, 443-446.
- DAVIDS, A., & OLIVER, G. R. Learning, performance, and retention under stress in emotionally disturbed and normal children. *J. Pers.*, 1960, 28, 368-381.
- DAVIS, R. H., & HARRINGTON, R. W. The effect of stimulus class on the problem-solving behavior of schizophrenics and normals. *J. abnorm. soc. Psychol.*, 1957, 54, 126-128.
- DIETHELM, O., & JONES, M. R. Influence of anxiety on attention, learning, retention, and thinking. *Arch. Neurol. Psychiat., Chicago*, 1947, 58, 325-336.
- EYSENCK, H. J. Conditioning and personality. *Brit. J. Psychol.*, 1962, 53, 299-305.
- FRANKS, C. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, 52, 143-150.
- GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, 20, 253-276.
- GILBERSTADT, H., & DAVENPORT, G. Some relationships between GSR conditioning and judgments of anxiety. *J. abnorm. soc. Psychol.*, 1960, 60, 441-443.
- GLADIS, M., & WISCHNER, G. J. Schizophrenic and normal response patterns to "aversive" and "neutral" associations in two paired-associate paradigms. *J. abnorm. soc. Psychol.*, 1962, 64, 249-256.
- HALL, K. R. L., & CROOKES, T. G. Studies in learning impairment: I. Schizophrenic and organic patients. *J. ment. Sci.*, 1951, 97, 725-737.
- HALL, K. R. L., & CROOKES, T. G. Studies in learning impairment: II. Psychoneurotic patients. *J. ment. Sci.*, 1952, 98, 273-279.
- HILBERSTAM, J. L. Some personality correlates

- of conditioning generalization and extinction. *Psychosom. Med.*, 1961, 23, 67-76.
- HOWE, E. S. GSR conditioning in anxiety states, normals, and chronic functional schizophrenic subjects. *J. abnorm. soc. Psychol.*, 1958, 56, 183-189.
- HULL, C. L. The formation and retention of associations among the insane. *Amer. J. Psychol.*, 1917, 28, 419-435.
- HUNT, W. L. Anxiety as drive and reactive inhibition. *J. gen. Psychol.*, 1959, 61, 159-167.
- HUSTON, P. E. Eye-hand co-ordination in schizophrenic patients and normals as measured by the pursuitmeter. *Psychol. Bull.*, 1932, 29, 662.
- HUSTON, P. E., & SHAKOW, D. Studies of motor function in schizophrenia: I. Speed of tapping. *J. gen. Psychol.*, 1936, 15, 63-106.
- HUSTON, P. E., & SHAKOW, D. Studies of motor function in schizophrenia: III. Steadiness. *J. gen. Psychol.*, 1946, 34, 119-126.
- HUSTON, P. E., & SHAKOW, D. Learning in schizophrenia: I. Pursuit learning. *J. Pers.*, 1948, 17, 52-74.
- HUSTON, P. E., & SHAKOW, D. Learning capacity in schizophrenia. *Amer. J. Psychiat.*, 1949, 105, 881-888.
- HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, 16, 39-82.
- KARRAS, A. The effects of reinforcement and arousal on the psychomotor performance of chronic schizophrenics. *J. abnorm. soc. Psychol.*, 1962, 65, 104-111.
- KENT, G. H. Experiments on habit formation in dementia praecox. *Psychol. Rev.*, 1911, 18, 375-410.
- KING, G. F., MERRELL, D. W., LOVINGER, E., & DENNY, M. R. Operant motor behavior in acute schizophrenics. *J. Pers.*, 1957, 25, 317-326.
- KING, H. E. *Psychomotor aspects of mental disease*. Cambridge: Harvard Univer. Press, 1954.
- KORCHIN, S. J., & LEVINE, S. Anxiety and verbal learning. *J. abnorm. soc. Psychol.*, 1957, 54, 234-240.
- LANG, P. J. The effect of aversive stimuli on reaction time in schizophrenia. *J. abnorm. soc. Psychol.*, 1959, 59, 263-268.
- LANG, P. J., & LUOTO, K. Mediation and associative facilitation in neurotic, psychotic, and normal subjects. *J. abnorm. soc. Psychol.*, 1962, 64, 113-120.
- LINDSLEY, O. R. Operant conditioning methods applied to research in chronic schizophrenics. *Psychiat. res. Rep.*, 1956, 5, 118-139.
- LONG, R. C. Praise and censure as motivating variables upon the motor behavior and learning of schizophrenics. Unpublished doctoral dissertation, University of Texas, 1960.
- LYKKEN, D. T. A study of anxiety in the sociopathic personality. *J. abnorm. soc. Psychol.*, 1957, 55, 6-10.
- LYNN, R. Russian theory and research on schizophrenia. *Psychol. Bull.*, 1963, 60, 486-498.
- MAILLOUX, N. M., & NEWBURGER, M. The work curves of psychotic individuals. *J. abnorm. soc. Psychol.*, 1941, 36, 110-114.
- MALMO, R. B., & AMSEL, A. Anxiety produced interference in serial rote learning, with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.*, 1948, 38, 440-454.
- MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *J. Pers.*, 1951, 19, 359-375.
- MALMO, R. B., & WALLERSTEIN, H. Rigidity and reactive inhibition. *J. abnorm. soc. Psychol.*, 1955, 50, 345-348.
- MEDNICK, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, 51, 536-542.
- MEDNICK, M. T., & LINDSLEY, O. R. Some clinical correlates of operant behavior. *J. abnorm. soc. Psychol.*, 1958, 57, 13-16.
- O'CONNOR, N. Reminiscence and work decrement in catatonic and paranoid schizophrenics. *Brit. J. med. Psychol.*, 1957, 30, 188-193.
- O'CONNOR, N., & RAWNSLEY, K. Two types of conditioning in psychotics and normals. *J. abnorm. soc. Psychol.*, 1959, 58, 157-161.
- PASCAL, G. R., & SWENSEN, C. Learning in mentally ill patients under conditions of unusual motivation. *J. Pers.*, 1952, 21, 240-249.
- PETERS, H. N. Circular pencil maze performance in chronic schizophrenics. *J. clin. Psychol.*, 1956, 12, 170-173.
- PETERS, H. N., & MUPHREE, O. D. The conditioned reflex in the chronic schizophrenic. *J. clin. Psychol.*, 1954, 10, 126-130.
- PAFFMAN, C., & SCHLOSBERG, H. The conditioned knee jerk in psychotic and normal individuals. *J. Psychol.*, 1936, 1, 201-205.
- RAFI, A. A. Motor performance of certain categories of mental patients. *Percept. mot. Skills*, 1960, 10, 39-42.
- ROBINSON, N. L. Paired-associate learning by schizophrenic subjects under conditions of

- personal and impersonal reward and punishment. Unpublished doctoral dissertation, Stanford University, 1958.
- RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, 97, 214-255.
- ROSENBAUM, G. Stimulus generalization as a function of clinical anxiety. *J. abnorm. soc. Psychol.*, 1956, 53, 281-285.
- SALZINGER, K., & PISONI, S. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, 57, 84-90.
- SALZINGER, K., & PISONI, S. Some parameters of the conditioning of verbal affect responses in schizophrenic subjects. *J. abnorm. soc. Psychol.*, 1961, 63, 511-516.
- SHAKOW, D. A study of certain aspects of motor co-ordination in schizophrenia with the prod meter. *Psychol. Bull.*, 1932, 29, 661.
- SHIPLEY, W. C. Further investigation of the preverbal tendency: Studies of catatonia. *Psychiat. Quart.*, 1934, 8, 736-744.
- SMITH, W. O. Rotary pursuit performance in reactive and process schizophrenics. Unpublished doctoral dissertation, Michigan State University, 1959.
- SPENCE, K. W., & TAYLOR, J. A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, 45, 265-272.
- TAYLOR, J. A., & SPENCE, K. W. Conditioning level in behavior disorders. *J. abnorm. soc. Psychol.*, 1954, 49, 497-502.
- TIZARD, J., & VENABLES, P. H. The influence of extraneous stimulation on the reaction time of schizophrenics. *Brit. J. Psychol.*, 1957, 48, 299-305.
- VENABLES, P. H. Factors in the motor behavior of functional psychotics. *J. abnorm. soc. Psychol.*, 1959, 58, 153-156.
- VENABLES, P. H., & TIZARD, J. The effect of stimulus light intensity on reaction time of schizophrenics. *Brit. J. Psychol.*, 1956, 47, 144-145. (a)
- VENABLES, P. H., & TIZARD, J. Paradoxical effects in the reaction time of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, 53, 220-224. (b)
- VENABLES, P. H., & TIZARD, J. Performance of functional psychotics on a repetitive task. *J. abnorm. soc. Psychol.*, 1956, 53, 23-26. (c)
- WELCH, L., & KUBIS, J. Conditioned PGR in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, 105, 372-381. (a)
- WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, 23, 83-91. (b)
- WINOKUR, G., GUZE, S., STEWART, M., PFEIFFER, E., STERN, J., & HORNING, F. Association of conditionability with degree of reactivity in psychiatric patients. *Science*, 1959, 129, 1423-1424.

(Received August 16, 1963)









